Predicting Financial Distress Using Multimodal Data: An Attentive and Regularized Deep Learning Method

Wanliu Che [a], Zhao Wang [a], Cuiqing Jiang [a, *], Mohammad Zoynul Abedin [b]

[a] School of Management, Hefei University of Technology, Hefei, Anhui, P.R. China

[b] School of Management, Swansea University, Bay Campus, Fabian Way, SA1 8EN Swansea, UK

Abstract

The proliferation of multimodal data provides a valuable repository of information for financial distress prediction. However, the use of multimodal data faces critical challenges, such as heterogeneity within and among modalities and difficulties in discriminating complementary and redundant information among modalities. To this end, we propose an attentive and regularized deep learning method for predicting financial distress using multimodal data, including financial indicators, current reports, and interfirm networks. Specifically, considering heterogeneity within and among modalities, we design three modality-specific attentions, i.e., ratio-aware, report-aware, and neighbor-aware attentions, for adaptively extracting key information from financial indicators, current reports, and interfirm networks, respectively. Considering difficulties in discriminating complementary and redundant information among modalities, we design a conditional entropy-based regularization to guide the method focusing on complementary information while discarding redundant information during modality fusion. We also propose the use of focal loss to address the class imbalance problem. Empirical evaluation shows that the proposed method significantly outperformed all benchmarked methods in terms of predictive and representation performance. We also provide key findings and implications for stakeholders.

Keywords: Financial distress prediction; multimodal data; deep learning; attention mechanism; conditional entropy

---

[*] Corresponding author. E-mail addresses: jiangcuiq@163.com; jiangcuiq2017@163.com (Cuiqing Jiang).

Wanliu Che and Zhao Wang are equal contributors to this work and designated as co-first authors.

1. Introduction

Financial distress refers to the inability of a company to meet its financial obligations due to insufficient revenues. Getting into financial distress affects a variety of different businesses with a severe negative impact on companies and may further lead to a variety of adverse consequences, such as investment loss, unemployment, and even government deficits (Wang et al., 2021). Financial distress prediction (FDP), as an effective tool to identify early warning signals of financial distress, is of great concern to company managers and investors in financial risk management.

The availability of multimodal data is reshaping the paradigm of FDP. Financial indicators, while playing a dominant role in FDP by reflecting a company's profitability, solvency, and liquidity (Xu et al., 2022; Zhang et al., 2022), only provide a snapshot of the company's financial position in the past year, which may not be sufficient to depict a comprehensive portrait (Kou et al., 2021). With the prevalence of digitalization technologies, multimodal data, exemplified by current reports (also known as 8-K filings) – a form of textual data, and interfirm networks – a form of network data, is becoming available online, providing important supplements to financial indicators. Current reports provide information on the occurrence of statutory material events of a company, enriching characterization of a company from the operational perspective (Jiang et al., 2022a). Interfirm networks present the influences from neighbor companies[1] that could continue on the target company, enriching characterization of a company from the relational perspective (Tobback et al., 2017). Financial indicators (numeric modality), current reports (textual modality), and interfirm networks (network modality) conjunctively and comprehensively characterize a company's financial status. Therefore, we focus on leveraging multimodal data for FDP in this study.

Predicting financial distress using multimodal data poses great challenges in terms of modality representation and modality fusion. For modality representation, feature utilities within modality are

[1] Neighbor companies refer to the companies that are connected to a target company within a relational network.

heterogeneous and such heterogeneity may even vary across observations, e.g., a same feature may present different utilities for predicting the financial distress of different companies. How to *explicitly and adaptively extract key information from each modality* is challenging. For modality fusion, multimodal data is generally complementary (i.e., unique information in each modality), but inevitably redundant (i.e., same information whereas in form of different modalities). For example, when a material corporate event occurs, like an acquisition, the deal terms and expected impacts are disclosed in detail in current reports. Meanwhile, these impacts may also be reflected in corresponding changes in financial indicators and interfirm networks. Complementary information contributes to performance improvement of FDP, whereas redundant information usually leads to adverse effects such as overfitting and predictive bias (Cui and Li, 2022). How to guide the modeling process to *focus on learning complementary information while discarding redundant information* is another challenge. Besides, the number of financially distressed companies is much less than that of normally operated companies in practice, and such class imbalance problem may adversely affect prediction performance.

To address these challenges, we propose a novel multimodal deep learning method, called attentive and regularized deep learning (ARDL), to leverage multimodal data for FDP. Specifically, to address the challenge of heterogeneous utilities of features in each modality, we design three attention modules (i.e., ratio-aware, report-aware, and neighbor-aware attentions) for generating effective representations for financial indicators, current reports, and interfirm networks, respectively. To address the challenge of information redundancy among modalities, we design a novel conditional entropy-based regularization to filter out the redundant information while maintaining complementary information during modality fusion based on the conditional entropy maximization principle. Besides, we propose the use of focal loss to address the notable class imbalance problem in FDP. With these design artifacts, ARDL could leverage multimodal data in a more meticulous manner, i.e., not only attentive to important information within modalities but also capable of refining unique information among modalities, leading to better predictive performance.

We have evaluated the proposed method using a multimodal dataset of Chinese companies in the

National Equities Exchange and Quotations (NEEQ) market. We compared ARDL with ten representative machine learning and deep learning methods. Empirical evaluation shows that ARDL significantly outperformed the benchmarked methods in terms of predictive performance, and its representation performance was superior to the benchmarked deep learning methods. The results also show that the combination of multiple modalities for FDP not always yields improved performance unless the key information within and among modalities is effectively extracted. Ablation study demonstrates that each design artifact of ARDL contributes to performance improvement. Interpretation analysis on attention weights illustrates how each feature within each modality differentially worked for FDP. Sensitivity analysis reveals how the regularization parameter effect the performance of ARDL. Robustness tests show that the effectiveness of ARDL is not affected by dataset selection and data diversity. We also provide the key findings and implications.

This study makes three contributions to the information processing and risk management domain. First, to our knowledge, this is the first study that integrates financial indicators, current reports, and interfirm networks information for predicting financial distress while considering the intra-modality importance and the inter-modality heterogeneity, as well as the complementarity and redundancy of modality information. We provide clear evidence that multimodal data provides more useful information for accurately predicting financial distress. We also uncover the existence of complementarity and redundancy among the information obtained from multiple modalities, both theoretically and empirically, providing the groundwork for future research. Second, we initialize a new way of multimodal data modeling by proposing ARDL. In contrast to the existing multimodal learning methods that primarily concentrate on identifying key information (e.g., using attentions) while overlooking the redundancy among modalities, ARDL introduces "extracting key information within modalities–extracting unique information among modalities–fusing information from multiple modalities" routine. This novel routine can explicitly extract essential and distinct knowledge from multimodal data. Third, we provide practically valuable insights for stakeholders that utilize multimodal data and AI tools for financial risk management. With the identified key factors from financial indicators, current reports, and interfirm networks, stakeholders could better understand the

signals indicating financial distress and takeover the AI tools.

The remainder of this paper is organized as follows. Section 2 reviews the relevant literature about financial distress and identifies the research gaps. Section 3 presents the details of our proposed ARDL model. We describe the empirical evaluation in Section 4 and report on the results in Section 5. Finally, we conclude our work in Section 6.

2. Literature Review

2.1 Predictors in Financial Distress Prediction

Financial distress prediction has drawn considerable attention from various communities such as finance and information science. Much literature has constructed multiple predictors for financial distress from the financial indicators, e.g., profitability, debt paying ability, and development ability. These financial indicators, which can directly provide insights into understanding the financial position of a company, are conventionally and continually used in FDP (Wang et al., 2021). Their effectiveness has been demonstrated by numerous studies (e.g., Li et al., 2021; Medina-Olivares et al., 2022). Nazareth and Ramana Reddy (2023) reviewed research on FDP and found that both previous and most existing studies relied solely on financial indicators as financial distress predictors.

With the flouring of digital technologies, various types of data (e.g., texts and networks) can be stored and accessed by companies, providing a variety of valuable sources of information for enhancing FDP performance. Text information reveals linguistic meaning through text and is universally regarded as a special type of repository for financial distress (Wang et al., 2021). Most studies have focused their efforts on company self-disclosure texts, such as annual reports, auditor reports, and current reports. These reports can provide valuable insights into a company's future performance (Wang et al., 2020). However, annual reports and auditor reports, which are often disclosed alongside the financial indicators, significantly overlap in content (Borchert et al., 2023). Current reports are another type of required disclosure document, notifying the public of important and emergency events occurring in the business's operations. Jiang et al. (2022a) extracted semantic features from current reports to predict financial distress and found that these features provide valuable clues about the existence of financial distress.

In addition to using textual information for FDP, another stream of research focuses on interfirm networks, i.e., exploiting the potential indicators from neighbor companies to enhance FDP performance. Long et al. (2022) constructed an interfirm network based on the relationship of sharing directors, supervisors, and senior management and demonstrated that incorporating relational information into credit risk assessment can improve predictive performance. Tobback et al. (2017) also found that integrating relational features is more effective in detecting high-risk companies, highlighting the significance of interfirm networks in FDP. In summary, the research shows financial indicators, current reports, and interfirm network data can all be effectively utilized for FDP.

2.2 Methods for Financial Distress Prediction

Both statistical and machine learning methods have been applied to FDP. Statistical methods were used earlier for building FDP models, such as discriminant analysis, logit regression analysis, and factor analysis, owing to their strong interpretability and easy-to-use characteristics (Iyer et al., 2016). Recently, machine learning models have been more widely used in FDP, including decision tree (DT) (Schmid et al., 2023), logistic regression (LR) (Dastile et al., 2020), support vector machines (SVM) (Sun et al., 2021), artificial neural networks (ANN) (Fu et al., 2020), and ensemble models, such as gradient boosting decision tree (GBDT) (Qian et al., 2022), eXtreme gradient boosting (XGB) (Xia et al., 2017) and LightGBM (Zhang et al., 2022). For instance, Geng et al. (2015) used DT, SVM, and ANN to develop FDP models based on 31 financial indicators, and found that ANN outperformed other models. Sun et al. (2021) enhanced SVM by integrating decomposition and fusion methods, leading to improved predictive performance in FDP.

Unfortunately, these statistical and machine learning methods face great challenges when dealing with high-dimension data with varying inherent properties, especially in the case of multimodal data. Deep learning methods, armed with their remarkable learning capability, have begun to catch up and provided great potential value in multimodal fusion. For example, Jiang et al. (2022b) introduced a deep financial distress prediction method to model corporate financial distress based on financial indicators and interfirm networks. Kraus and Feuerriegel (2017) combined both financial indicators and current reports with deep neural networks and obtained an integrated multimodal representation for financial decision support,

achieving outstanding performance. Inspired by these efforts, we aim to design a deep learning model for effectively leveraging multimodal data for FDP.

Besides, class imbalance has always been an inevitable problem that may jeopardize the prediction performance (Sun et al., 2020). However, previous studies on FDP either directly ignored the impact of class imbalance or constructed balanced samples (e.g., through oversampling or undersampling) (Nazareth and Ramana Reddy, 2023). The former cannot reflect the real status of the financial market, while the latter damages the original data distribution. This has motivated us to address the class imbalance problem in FDP while preserving the original data distribution.

2.3 Multimodal Modeling

Multimodal modeling commonly involves two main components: modality representation and modality fusion. Modality representation refers to the transformation of raw data input from each modality into a dense vector representation that can characterize the intrinsic attributes and patterns. For example, Wang et al. (2020) separately modeled financial and textual modalities to represent original data for FDP. The financial representation was constructed using the numeric financial ratios in a well-structure format, while textual representation was generated using a combination of lexicon-based techniques and the bag of words method. Beaver et al. (2019) generated a network representation by constructing an interfirm network based on group affiliation data and using financial ratios as financial representation for default prediction. Although these studies treated each modality as an independent encoding, most of them ignored the intra-modal importance, which is a key factor affecting the final prediction performance. In FDP, Matin et al. (2019) employed the attention mechanisms on textual modality by endowing different attentions at the word and sentence level, enhancing the predictive ability and interpretability of textual modality. However, they only applied attention mechanisms on unimodal data, ignoring the heterogeneity among multiple modalities. Besides, no prior studies have focused on the modal-specific attention mechanisms in FDP, which motivates us to design modal-specific attention mechanisms, not only detecting companies that will be at risk but also helping to understand why companies will be at risk.

Following the modality fusion paradigm, multimodal methods are mainly divided into two categories: early fusion (feature-level fusion) and late fusion (decision-level fusion) (LeCun et al., 2015). Early fusion-based methods extract multimodal features using specific subnetworks and then perform fusion at the feature level through various techniques such as concatenation, summation, and attention mechanisms. For example, Borchert et al. (2023) achieved multimodal fusion by concatenating the financial and textual features extracted by the specific subnetworks, whereas Yang et al. (2021) used the fusion technique of summation to combine the textual and visual features. These techniques are easy to use but may yield unstable results (Lu et al., 2023; Yang et al., 2023). Modality fusion based on attention mechanisms assigns weights to features through their distinct importance and can further interpret certain feature impacts by visualizing attention weights. For example, the attention-fused network (e.g., RCMA, Wang et al., 2023 and NIGCM, Jiang, et al., 2022b) fuse financial and textual (or network) features extracted by specific subnetworks by designing different attention-fused blocks, yielding improved predictive performance and more interpretability. Late fusion-based methods independently predict a result for each modality and then combine them using techniques such as weighing or voting. In the late fusion methods, multimodal fusion representation is limited by the absence of cross-modality interactions among the data. Although the early fusion methods are by far the most commonly used, they either ignore the redundant information or simultaneously diminish the complementary and redundant effects. This limitation has inspired us to develop a novel multimodal fusion method that not only filter out the redundant information but also retain complementary information during modality fusion, thus enhancing the performance of multimodal learning in FDP.

2.4 Research Gaps

Multimodal data-based FDP has gained significant attention, and Table 1 summarizes representative studies in terms of the use of modalities, modality representation, and modality fusion. From the perspective of modality, while previous studies have mostly focused on using either textual modality or network modality for FDP, as a supplement to financial indicators (i.e., numeric modality), no research has yet been dedicated to exploring the combined power of numeric, textual, and network modalities for FDP. Moreover,

utilizing these modalities entails several challenges, such as intra-modality importance and inter-modality heterogeneity, as well as the complementarity and redundancy of modality information as mentioned in Table 1. From the perspective of multimodal modeling, modality representation and modality fusion are crucial components that drive performance. For modality representation, while previous studies have focused on unimodal or bimodal representation, how to effectively generate representations considering both the inter-modal heterogeneity and intra-modal importance is still challenging. For modality fusion, attention, with its ability to adaptively identify key information during modality fusion, has always been the go-to approach. However, the attention mechanism simultaneously amplifies (diminishes) the complementary and redundant effects, e.g., a high attention weight intensifies the presence of both discriminative information and useless information. How to guide the modeling process in a way of focusing on learning complementary information while discarding redundant information is also challenging. We strive to bridge these gaps by proposing an attentive and regularized deep learning method.

Table 1. Representative Studies on FDP Using Multimodal Data.

Study	Use of numeric modality	Use of textual modality	Use of network modality	Modality representation		Modality fusion	
				Inter-modal heterogeneity	Intra-modal importance	Information complementary	Information redundancy
Geng et al. (2015)	√						
Chen et al. (2016)	√						
Tobback et al. (2017)	√		√			√	
Matin et al. (2019)	√	√			√	√	
Beaver et al. (2019)	√		√	√		√	
Yıldırım et al. (2021)	√		√			√	
Li et al. (2021)	√	√		√		√	
Lee et al. (2021)	√		√			√	
Jiang et al. (2022a)	√	√				√	
Wang et al. (2023)	√	√		√	√	√	
This study	√	√	√	√	√	√	√

3. Proposed Method

Given the immense potential of multimodal data, represented by financial indicators, current reports, and interfirm networks, for improving the performance of FDP, and the lack of a promising solution for effectively leveraging multimodal data in terms of tackling heterogeneity within and among modalities and tradeoff between complementarity and redundancy among modalities, we propose an attentive and regularized deep learning method (ARDL) for FDP. The idea behind ARDL lies in bringing financial indicators, current reports, and interfirm networks into a unified deep learning framework following the "extracting key information within modalities–extracting unique information among modalities–fusing information from multiple modalities" routine. With the tailored components, namely modality-specific attentions and conditional entropy-based regularization (will be discussed later), ARDL allows to explicitly and adaptively extract key information within each modality and alleviate information redundancy during multimodal fusion, contributing to prediction performance improvement.

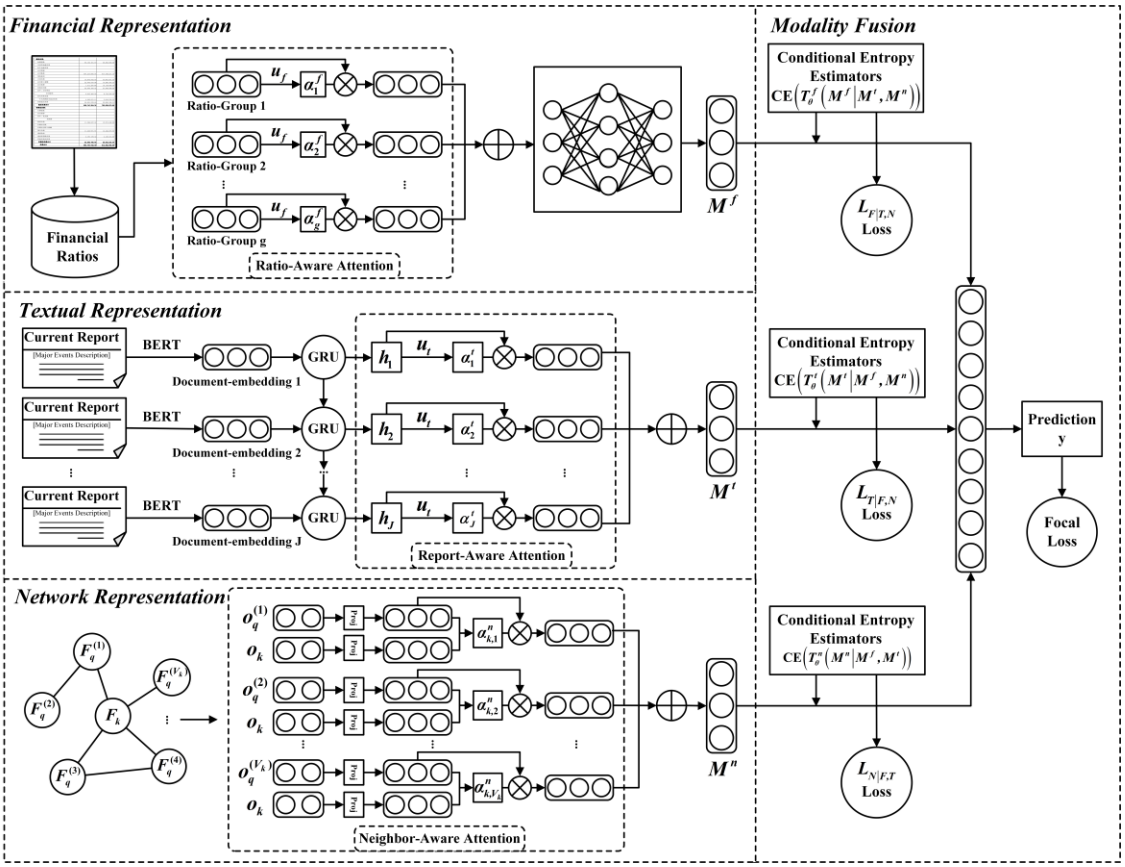

Figure 1. Framework of ARDL.

3.1. Overview of ARDL

Figure 1 illustrates the framework of ARDL. Given the multimodal data, ARDL consists of three parts: modality representation, modality fusion, and risk prediction. In the part of modality representation, we design three modality-specific attention modules, i.e., ratio-aware, report-aware, and neighbor-aware attentions, to generate effective representations for financial indicators, current reports, and interfirm networks, respectively. In the part of modality fusion, we design a novel conditional entropy-based regularization and incorporate it into the loss function of ARDL. By minimizing this regularization, redundant information among modalities can be effectively reduced, while simultaneously encouraging the preservation of unique information from different modalities. In the part of risk prediction, considering the class imbalance problem (i.e., financially distressed companies account for a small proportion), we introduce the focal loss (Lin et al., 2020) to the loss function.

3.2. Attentive Modality Representation

For financial representation, given that different financial ratios naturally reflect various aspects of financial position (e.g., profitability, solvency, development capability, operational capabilities, and finance structure), we categorized them into distinct ratio groups to generate a joint effect for prediction. Table 2 summarizes five different types of ratio groups along with their representative ratios in FDP. Obviously, the contribution of different ratio groups varies for FDP, and the distinct predictive power of each ratio group need to be emphasized. Hence, the ratio-aware attention is designed to explicitly identify the distinct importance of each ratio group. Unlike traditional attention mechanisms that are directly applied to individual ratios, ratio-aware attention is applied at the group level, assigning a distinct attention weight to each group. This helps to reduce the impact of individual less informative ratios. Furthermore, by grouping ratios based on their financial meaning, ratio-aware attention facilitates incorporating domain knowledge and provides a clearer organizational structure, thereby enhancing the model's interpretability.

Table 2. Ratio Groups in Financial Indicators.

Category	Contents	Ratio Group
Profitability	Return on assets, Return on equity, Net profit ratio, Net profit to current asset, Net profit to fixed asset, Ebit to asset	G1

Solvency	Current ratio, Quick ratio, Asset liability ratio, Debt equity ratio, Debt tangible equity ratio, Current liability coverage	G2
Development capacity	Operating revenue growth rate, Net profit growth rate, Assets growth rate, Net operating cash flow growth rate, Operation cash per share growth rate, Equity growth rate	G3
Operational capabilities	Inventory turning rate, Receivable turnover ratio, Accrued payable rate, Equity rate, Net operating cycle, Working capital total rate	G4
Finance structure	Current asset ratio, Fixed asset ratio, Equity to fixed asset ratio, Current liability ratio, Equity ratio, Working capital to equity	G5

Ratio-aware attention for financial representation. We represent the ratio groups as $\boldsymbol{R} = [\boldsymbol{r}_1, \boldsymbol{r}_2, \dots, \boldsymbol{r}_g]$, where g refers to the total number of ratio groups. To emphasize their distinct contributions, we incorporate an attention mechanism into the ratio groups. Attention is a prevalent weight learning scheme aimed to assign and optimize a set of weights that correspond to the input elements, endowing neural networks with the ability to focus more on the salient elements (Dikmen and Burns, 2022). The general process of attention mechanism can be described as mapping a query and a set of key-value pairs to an output, in which the query, key, and value are represented as vectors (Vaswani et al., 2017). For each input ratio group, ratio-aware attention assigns it a trainable query vector $\boldsymbol{u}_f$, aimed at capturing the relevance between the current ratio group and the other groups. The same ratio group acts as the key-value pairs to enhance interpretability (Yang et al., 2021). During the training process, each ratio group is matched with the query vector to produce the importance weight α^f via a softmax function:

$$\alpha_i^f = \frac{\exp\left(\boldsymbol{u}_f^{\mathrm{T}} \boldsymbol{r}_i\right)}{\sum_{i=1}^{g} \exp\left(\boldsymbol{u}_f^{\mathrm{T}} \boldsymbol{r}_i\right)} \tag{1}$$

where $\boldsymbol{u}_f$ is initially randomized and subsequently optimized during the training process.

The financial representation $\boldsymbol{M}^f = \left\{M_1^f, M_2^f, \dots, M_{d_f}^f\right\}$, with the representation dimensionality at d_f, is then generated using merged ratio group and a multilayer perceptron (MLP) as follows:

$$\boldsymbol{M}^f = \mathrm{MLP}\left(\sum\nolimits_{i=1}^{g} \alpha_i^f \boldsymbol{r}_i\right) \tag{2}$$

where MLP(·) denotes an MLP layer. While directly using the merged ratio group as the financial

representation seems straightforward, the underlying predictive signals within the ratio groups would remain untapped without deeper processing. By applying an MLP encoder on the merged ratio group, we enable the model to derive a more informative and diversified representation. The benefits stem from MLP's exceptional capacity to capture intricate nonlinear patterns and multidimensional relationships embedded within the structured data (Gorishniy et al., 2022).

For textual representation, considering that current reports encompass diverse descriptions of material events with various semantic and syntactic relationships and are presented in a sequential manner (Jiang et al., 2022a), we initially employ a pre-trained language model named BERT (i.e., Bidirectional Encoder Representations from Transformers), which has achieved remarkable success in natural language processing (Devlin et al., 2018), to get document embeddings. Then, recognizing that these document embeddings inherently maintain a sequential structure, we adopt a Gate Recurrent Unit (GRU) network. GRU possesses a distinctive advantage in handling sequential data, enabling it to capture the temporal relationships among reports for more in-depth representations (Fan and Ilk, 2020). The GRU network can effectively accommodate time dependencies of document embeddings, ensuring that the information from the previous sequence is retained when integrating a series of sequences for prediction. However, since each report reflects a distinctive material event and contributes differently to the prediction, relying solely on GRU is insufficient. Hence, we design the report-aware attention on current reports, which incorporates an attention mechanism into GRU to explicitly emphasize important reports, and thus generate the final textual representation. The processes are detailed as follows.

To extract document embedding vectors of the current reports, we map each report into BERT and derive its document embedding from the final hidden state of the [CLS] token in BERT, thereby obtaining the embedding matrix $\boldsymbol{E} = [\boldsymbol{e}_1, \boldsymbol{e}_2, \dots, \boldsymbol{e}_J] \in \mathbb{R}^{d_e \times J}$ for all current reports of a company, where d_e is the dimensionality of the BERT embeddings and J is the total number of reports.

Report-aware attention for textual representation. Given the obtained document embedding of each report, GRU is employed to extract deep representations by exploring the temporal relationships

among all the document embeddings. A GRU unit comprises two gates (reset and update gates) and a hidden state. These gates regulate the flow of information through the unit. The reset gate determines which information from the previous timestep should be forgotten, and the update gate decides how much of the new input should be incorporated. The hidden state captures the current state of the unit. The detailed processes of these two gates are as follows:

$$\boldsymbol{r}_j = \text{sigmoid}\left(\boldsymbol{W}_r \cdot \left[\boldsymbol{h}_{j-1}, \boldsymbol{e}_j\right] + \boldsymbol{b}_r\right) \tag{3}$$

$$\boldsymbol{z}_j = \text{sigmoid}\left(\boldsymbol{W}_z \cdot \left[\boldsymbol{h}_{j-1}, \boldsymbol{e}_j\right] + \boldsymbol{b}_z\right) \tag{4}$$

$$\widehat{\boldsymbol{h}}_j = \tanh\left(\boldsymbol{W}_h \cdot \left[\boldsymbol{r}_j \otimes \boldsymbol{h}_{j-1}, \boldsymbol{e}_j\right] + \boldsymbol{b}_h\right) \tag{5}$$

$$\boldsymbol{h}_j = \left(1 - \boldsymbol{z}_j\right) \otimes \boldsymbol{h}_{j-1} + \boldsymbol{z}_j \otimes \widehat{\boldsymbol{h}}_j \tag{6}$$

where $\boldsymbol{W}_r$, $\boldsymbol{W}_z$, and $\boldsymbol{W}_h$ are the weight matrices of reset gate, update gate, and hidden state, respectively. $\boldsymbol{b}_r$, $\boldsymbol{b}_z$, and $\boldsymbol{b}_h$ are the corresponding bias vectors. $\boldsymbol{h}_j$ is the j-th output state of the hidden layers of GRU. By taking these output states from each timestep of the GRU, we obtain the textual embedding $\boldsymbol{H} = \left[\boldsymbol{h}_1, \boldsymbol{h}_2, \ldots, \boldsymbol{h}_J\right] \in \mathbb{R}^{d_h \times J}$, where d_h is the dimensionality of the textual embedding.

Next, an attention mechanism is introduced to highlight these reports that are informative in revealing financial distress. Compared with traditional attention mechanisms applied in the GRU layer, report-aware attention maintains a similar structural setup. However, the key innovation lies in its direct application to the output of the GRU hidden states, eliminating the requirement for transforming the initial output. This approach offers the advantage of better capturing the correlations and temporal patterns of different time steps within the report sequence, while also reducing model parameters and computational complexity. Specifically, for each output state of the current reports, report-aware attention assigns it a trainable query vector $\boldsymbol{u}_t$ to capture the relevance between the current state and other states in the sequence. The same output state acts as the key-value pairs to improve interpretability. During the training process, each state is matched with the query vector to produce the importance weight α^t via a softmax function:

$$\alpha_j^t = \frac{\exp\left(\boldsymbol{u}_t^{\mathrm{T}} \boldsymbol{h}_j\right)}{\sum_{j=1}^{J} \exp\left(\boldsymbol{u}_t^{\mathrm{T}} \boldsymbol{h}_j\right)} \tag{7}$$

The textual representation $\boldsymbol{M}^t = \{M_1^t, M_2^t, \ldots, M_{d_t}^t\}$ for all current reports of a company can be obtained by a weighted sum of the output states of the GRU in Eq. (8), where d_t refers to the dimensionality of the textual representation.

$$\boldsymbol{M}^t = \sum_{j=1}^{J} \alpha_j^t \boldsymbol{h}_j \tag{8}$$

Interfirm networks are a type of relational graph characterized by various relationships and interdependencies between companies (Bi et al., 2022; Topuz et al., 2021). As for representing networks, since different neighbor companies have distinct attribute information, directly aggregating neighbor information without considering their individualized impacts inevitably weakens the contributions of the more influential neighbors to the prediction. Hence, we design the neighbor-aware attention to adaptively aggregate information from neighboring nodes, emphasizing those with a stronger influence for the network representation. The processes are detailed as follows.

To quantify the influence of a company's immediate neighbors, our method first generates an undirected graph $G = (A, S)$ to model the interdependent relationships among companies. In G, the variable $A = K \cup Q$ represents the node set involving K and Q separately, as the set of target companies and their corresponding neighbor companies. The variable $S = \{(F_k, F_q) | F_k \in K, F_q \in Q\}$ is the edge set, each of which represents the connection between a target company F_k and its neighbor company F_q. We denote $\boldsymbol{o}_k$ and $\boldsymbol{o}_q$ as the original node attributes of F_k and F_q, respectively. Notably, the target company F_k may be connected to multiple neighbor companies, denoted as $Q_k = \{F_q^{(1)}, F_q^{(2)}, \ldots, F_q^{(V_k)}\}$, with the corresponding attribute set $\boldsymbol{O}_q = \{\boldsymbol{o}_q^{(1)}, \boldsymbol{o}_q^{(2)}, \ldots, \boldsymbol{o}_q^{(V_k)}\}$, where V_k is the number of neighbor companies of F_k. According to the empirical findings of Long et al. (2022), shared directors, supervisors, and senior management (DSS) relationships between companies have been identified as the most effective for constructing interfirm networks when evaluating credit risk. Accordingly, the interfirm networks in this study were constructed based on the DSS relationships.

Neighbor-aware attention for network representation. Given the interfirm networks G, traditional deep neural networks encounter significant challenges in capturing the intricate relationships between nodes and edges inherent in this type of relational data. Therefore, we employ graph neural network (GNN) as a solution. GNN offers a distinctive advantage in modeling both node attributes and connectivity information, enabling the network to learn the interaction patterns among companies. Within each GNN layer, two crucial functional modules exist: information aggregation (local transition function) and information update (local output function). The information aggregation module aggregates attribute information from neighboring nodes, while the information update module combines the aggregated information and previous node features to generate new representations for each node. However, each neighbor contains distinct attribute information, and relying solely on GNN is insufficient for establishing satisfactory node representations. One key reason is that GNN-based representation learning lacks explicit modeling of distinct neighbors (Ye and Ji, 2021). Hence, we design the neighbor-aware attention in interfirm networks, which incorporates the attention mechanism into GNN to explicitly learn the distinct contribution of each neighbor.

Unlike traditional attention mechanisms applied in graphs, neighbor-aware attention introduces a novel operation by transposing the embeddings of target nodes and performing element-wise multiplication with the tanh-activated embeddings of neighbors. This operation facilitates a more effective understanding of the topological structure of each node's neighborhood and the distribution of node attributes within that neighborhood. As a result, neighbor-aware attention enhances the overall capability of the model to capture intricate relationships and structural patterns in the relational data. Considering the target company F_k and its v-th neighbor $F_q^{(v)}$, the detailed process is as follows:

$$C_{k,v}\left(F_k, F_q^{(v)}\right) = (\boldsymbol{W}_k \boldsymbol{o}_k)^{\mathrm{T}} \tanh\left(\boldsymbol{W}_q^{(v)} \boldsymbol{o}_q^{(v)}\right) \tag{9}$$

where tanh is a nonlinear activation function, and $\boldsymbol{W}_k$ and $\boldsymbol{W}_q^{(v)}$ are learnable parameters that project the original attributes, $\boldsymbol{o}_k$ and $\boldsymbol{o}_q^{(v)}$, into high-dimensional feature spaces through linear transformation. We use the inner product here to compute the attention weights, which reflects the relevance between the two nodes

(Vaswani et al., 2017). Thereafter, the softmax function is employed to normalize the attention weights across all neighbors, expressed as follows:

$$\alpha_{k,v}^{n} = \frac{\exp\left(C_{k,v}\left(F_k, F_q^{(v)}\right)\right)}{\sum_{v=1}^{V_k}\exp\left(C_{k,v}\left(F_k, F_q^{(v)}\right)\right)} \tag{10}$$

The network representation for the target company F_k is then formalized by aggregating information from its weighted neighbors as follows:

$$\boldsymbol{M}^{n} = \sum_{v=1}^{V_k} \alpha_{k,v}^{n}\, \boldsymbol{W}_q^{(v)} \boldsymbol{o}_q^{(v)} \tag{11}$$

Through the above process, we obtain $\boldsymbol{M}^{n} = \{M_1^n, M_2^n, \ldots, M_{d_n}^n\}$ as the network representation that explicitly aggregates all neighbors' information in a differential manner, where d_n refers to the dimensionality of the network representation.

3.3. Conditional Entropy-Based Regularization

Having obtained the attentive representations for each modality using the modality-specific attentions, we integrate them into a unified multimodal representation for FDP. Common fusion strategies fall into two types: simple fusion (e.g., concatenation or summation) and attention-based fusion. Simple fusion, however, does not consider both complementarity and redundancy among modalities (Huang et al., 2021). Although attention-based fusion assigns modalities different weights according to their importance, it simultaneously amplifies or diminishes the complementary and redundant effects based on these weights. Hence, we design a conditional entropy-based regularization during modality fusion and incorporate it into the loss function of the main network. By minimizing this regularization, redundant information among modalities can be effectively reduced and unique information in each modality can be maintained.

Conditional entropy measures the uncertainty of one random variable given another (Shannon, 1948). It quantifies the additional information provided by one variable when the other is known. A higher conditional entropy indicates a greater amount of unique information and less redundancy between the variables. This naturally leads us to incorporate conditional entropy-guided knowledge into the learning

process. By maximizing the conditional entropy of the modality representations, we can minimize redundant information within these representations. However, estimating conditional entropy is challenging, as the underlying probability distributions are unavailable in practice. Commonly used methods for estimating conditional entropy include binning, the nearest-neighbor estimator, and the kernel-based estimator (Pichler et al., 2022). Nevertheless, these methods tend to become intractable in high-dimensional scenarios and are incompatible with gradient descent-based deep learning methods. Recent works propose optimizing the variational bounds of conditional entropy by training neural networks (e.g., Liu et al., 2020; Ge et al., 2023), which have shown impressive performance in conditional entropy estimation. Hence, we estimate conditional entropy by optimizing its upper-bound with cross-entropy (Shalev et al., 2022). Through optimizing the upper-bound of the conditional entropy across the three modality representations, we can maximize their conditional entropy, thereby enhancing the generation of complementary representations during modality fusion.

Taking the financial representation $\boldsymbol{M}^f$ as an example, we elaborate on the estimation of the conditional entropy $H\left(\boldsymbol{M}^f \middle| \boldsymbol{M}^t, \boldsymbol{M}^n\right)$.

Conditional entropy estimation. The inputs to the conditional entropy estimator include the financial representation $\boldsymbol{M}^f$ and the other two modality representations, $\boldsymbol{M}^t$ and $\boldsymbol{M}^n$. Note that $\boldsymbol{M}^f$, $\boldsymbol{M}^t$, and $\boldsymbol{M}^n$ do not need to have the same feature dimensions. Eq. (12) formulates the upper-bound for the conditional entropy of financial representation, expressed as $\mathrm{CE}(T_\theta^f(\boldsymbol{M}^f|\boldsymbol{M}^t, \boldsymbol{M}^n))$. The theoretical proof of this upper-bound is available in Appendix A.

$$H(\boldsymbol{M}^f|\boldsymbol{M}^t, \boldsymbol{M}^n) \leq \mathrm{CE}(T_\theta^f(\boldsymbol{M}^f|\boldsymbol{M}^t, \boldsymbol{M}^n)) \tag{12}$$

where the cross-entropy term $\mathrm{CE}(T_\theta^f(\boldsymbol{M}^f|\boldsymbol{M}^t, \boldsymbol{M}^n)) = -\mathbb{E}_{P(\boldsymbol{M}^f, \boldsymbol{M}^t, \boldsymbol{M}^n)} \log(T_\theta^f(\boldsymbol{M}^f|\boldsymbol{M}^t, \boldsymbol{M}^n))$. $T_\theta^f(\cdot)$ is a neural network parameterized with θ, trained to approximate the conditional distribution $P\left(\boldsymbol{M}^f \middle| \boldsymbol{M}^t, \boldsymbol{M}^n\right)$.

While the upper-bound of the conditional entropy has strong consistency (Shalev et al., 2022), directly estimating the conditional entropy by minimizing the cross-entropy remains challenging, as it requires the prior knowledge of $P\left(\boldsymbol{M}^f, \boldsymbol{M}^t, \boldsymbol{M}^n\right)$, which is difficult to obtain in practice. To address this, we employ the

entropy chain rule to achieve a more accurate estimation. This approach allows us to represent the conditional entropy $H\left(\boldsymbol{M}^f \middle| \boldsymbol{M}^t, \boldsymbol{M}^n\right)$ as the sum of a sequence of conditional entropies in Eq. (13). It facilitates the training of neural networks in a self-supervised manner and does not require any prior knowledge about the underlying distributions of $\boldsymbol{M}^f$, $\boldsymbol{M}^t$, or $\boldsymbol{M}^n$.

$$H(\boldsymbol{M}^f|\boldsymbol{M}^t, \boldsymbol{M}^n) = \sum_{d=1}^{d_f} H\left(M_d^f \middle| \boldsymbol{M}^t, \boldsymbol{M}^n, \boldsymbol{M}_{d-1}^f\right) \tag{13}$$

where M_d^f represents the d-th feature in $\boldsymbol{M}^f$, $\boldsymbol{M}_{d-1}^f$ represents the first $d-1$ features in $\boldsymbol{M}^f$, and $H\left(M_1^f \middle| \boldsymbol{M}^t, \boldsymbol{M}^n, \boldsymbol{M}_0^f\right)$ abbreviates $H\left(M_1^f \middle| \boldsymbol{M}^t, \boldsymbol{M}^n\right)$.

Building on this idea, we construct a total of d_f conditional entropy estimators, each represented by a neural network $T_{\theta_d}^f$, which is comprised of several fully connected layers with parameters θ_d. The inputs to the neural networks are composed of the first $d-1$ features in $\boldsymbol{M}^f$, along with $\boldsymbol{M}^t$ and $\boldsymbol{M}^n$. The output is the estimated conditional entropy of the d-th feature in $\boldsymbol{M}^f$.

However, directly estimating the probability distribution of the d-th feature in $\boldsymbol{M}^f$ to infer its conditional entropy is challenging due to its unbounded value range. A common solution is to discretize continuous features into discrete classes, enhancing the effectiveness of probability distribution estimation (González-López et al., 2020). Specifically, within each minibatch, we initially sort the d-th feature values of the financial representations and partition them into predefined bins. These bins, representing intervals that cover specific value ranges, can effectively discretize the continuous values. These feature values are then transformed into multiple integer classes based on their assigned bins. Finally, one-hot encoding is applied to each integer class, converting the integers into a binary representation as the estimated label.

After the discretization process, the estimated conditional entropy of the d-th feature in $\boldsymbol{M}^f$, denoted as $\widehat{H}\left(M_d^f \middle| \boldsymbol{M}^t, \boldsymbol{M}^n, \boldsymbol{M}_{d-1}^f\right)$, can be derived by minimizing the cross-entropy:

$$\widehat{H}_\Theta\left(M_d^f \middle| \boldsymbol{M}^t, \boldsymbol{M}^n, \boldsymbol{M}_{d-1}^f\right) = \inf_{\theta_d \in \Theta} \mathrm{CE}\left(T_{\theta_d}^f\left(M_d^f|\boldsymbol{M}^t, \boldsymbol{M}^n, \boldsymbol{M}_{d-1}^f\right)\right) \tag{14}$$

Minimizing the cross-entropy using neural networks allows us to optimize the upper-bound of conditional entropy. Given a minibatch samples of size B, the empirical estimator for this cross-entropy is expressed as:

$$\widehat{\mathrm{CE}}(T_{\theta_d}^f(M_d^f|\boldsymbol{M}^t,\boldsymbol{M}^n,\boldsymbol{M}_{d-1}^f)) = -\frac{1}{B}\sum_{i=1}^{B}\log\left(T_{\theta_d}^f(M_{i,d}^f|\boldsymbol{M}_i^t,\boldsymbol{M}_i^n,\boldsymbol{M}_{i,d-1}^f)\right) \tag{15}$$

where $M_{i,d}^f$, $\boldsymbol{M}_i^t$, $\boldsymbol{M}_i^n$, and $\boldsymbol{M}_{i,d-1}^f$ individually symbolize the M_d^f, $\boldsymbol{M}^t$, $\boldsymbol{M}^n$, and $\boldsymbol{M}_{d-1}^f$ of the i-th sample.

Therefore, the estimated conditional entropy of the financial representation, $\widehat{H}(\boldsymbol{M}^f|\boldsymbol{M}^t,\boldsymbol{M}^n)$, can be obtained by summing these estimated conditional entropies $\widehat{H}(M_d^f|\boldsymbol{M}^t,\boldsymbol{M}^n,\boldsymbol{M}_{d-1}^f)$, for $d = 1,\dots,d_f$. Similarly, the estimated conditional entropy of textual and network representations can also be obtained using the same estimation scheme, that is $\widehat{H}(\boldsymbol{M}^t|\boldsymbol{M}^f,\boldsymbol{M}^n)$ and $\widehat{H}(\boldsymbol{M}^n|\boldsymbol{M}^f,\boldsymbol{M}^t)$.

Conditional entropy-based loss function. The conditional entropy-based loss function consists of three parts: $L_{F|T,N}$, $L_{T|F,N}$, and $L_{N|F,T}$. Formally, the conditional entropy-based loss for financial representation $L_{F|T,N}$ can be expressed as follows:

$$L_{F|T,N} = -\frac{1}{B}\sum_{i=1}^{B}\sum_{d=1}^{d_f}\log\left(T_{\theta_d}^f(M_{i,d}^f|\boldsymbol{M}_i^t,\boldsymbol{M}_i^n,\boldsymbol{M}_{i,d-1}^f)\right) \tag{16}$$

Similarly, the conditional entropy-based losses for the textual and network representations, denoted as $L_{T|F,N}$ and $L_{N|F,T}$, are expressed as follows:

$$L_{T|F,N} = -\frac{1}{B}\sum_{i=1}^{B}\sum_{d=1}^{d_t}\log\left(T_{\theta_d}^t(M_{i,d}^t|\boldsymbol{M}_i^f,\boldsymbol{M}_i^n,\boldsymbol{M}_{i,d-1}^t)\right) \tag{17}$$

$$L_{N|F,T} = -\frac{1}{B}\sum_{i=1}^{B}\sum_{d=1}^{d_n}\log\left(T_{\theta_d}^n(M_{i,d}^n|\boldsymbol{M}_i^f,\boldsymbol{M}_i^t,\boldsymbol{M}_{i,d-1}^n)\right) \tag{18}$$

Therefore, the conditional entropy-based loss L_{fus} for the entire multimodal fusion module is computed as the average of $L_{F|T,N}$, $L_{T|F,N}$, and $L_{N|F,T}$:

$$L_{fus} = (L_{F|T,N} + L_{T|F,N} + L_{N|F,T})/3 \tag{19}$$

By incorporating the conditional entropy-based loss L_{fus} into the loss function of the main network, we introduce a conditional entropy-based regularization term that guides the model to maximize the

conditional entropy of modality representations during modality fusion, thereby alleviating redundant information. During the training process, the regularization term iteratively optimizes the parameters of the conditional entropy estimators using gradient descent: $\theta_d \leftarrow \theta_d + \widetilde{\nabla} L_{fus}(\theta_d)$. As the conditional entropy estimators and the main network are trained together, the estimated loss produced by the conditional entropy estimators simultaneously adjusts the parameters (i.e., weight and bias) of the main network, toward the generation of modality representations with lower redundancy and greater uniqueness. Through iterative parameters optimization to minimize the loss function, the generation of modality representations is iteratively enhanced, not only effectively reducing redundant information among modalities, but also encouraging the preservation of unique information from different modalities. These enhanced modality representations thereby conjunctively contribute to prediction performance improvement. We obtain the fused multimodal representation $\boldsymbol{M}^{comb}$ by concatenating these modality representations.

3.4. Learning Strategy of ARDL

Given the substantial challenge of class imbalance in FDP modeling, we address this issue by incorporating the focal loss function (Lin et al., 2020), which is a modification of the standard cross-entropy loss. Unlike commonly used imbalanced processing methods, for example, oversampling and undersampling, which manipulate the original data distribution before training, focal loss addresses this issue by adjusting sample weights without distribution change. It can down-weight the influence of easy-to-clarify samples during training, and gives more attention to hard-to-clarify samples by introducing hyperparameters (Lin et al., 2020). By this means, the contribution of hard-to-clarify samples to the loss function is enhanced, thereby alleviating the class imbalance problem.

Subsequently, we employ a fully connected layer with a sigmoid activation function to generate the final prediction. The computational formula is expressed as follows:

$$\hat{y} = \text{sigmoid}\left(\boldsymbol{W}^{comb}\boldsymbol{M}^{comb} + \boldsymbol{b}^{comb}\right) \tag{20}$$

where $\hat{y}$ is the predicted probability of the model for the corresponding class label y. $\boldsymbol{W}^{comb}$ and $\boldsymbol{b}^{comb}$ are the trainable parameters. Accordingly, given the two types of the loss functions, i.e., the focal loss L_{focal}

and the conditional entropy loss L_{fus}, the overall loss L_{ARDL} is expressed as:

$$L_{ARDL} = L_{focal} + \lambda L_{fus} \tag{21}$$

where λ is a tunable parameter greater than 0 that controls the degree of redundancy during modality fusion.

In a minibatch setting of size B, the formal description of the focal loss function is expressed as follows:

$$L_{focal} = -\sum_{i=1}^{B} (\beta(1-\hat{y}_i)^{\gamma} y_i \ln(\hat{y}_i) + (1-\beta)(\hat{y}_i)^{\gamma}(1-y_i)\ln(1-\hat{y}_i)) \tag{22}$$

where β is the parameter that balances the distribution of positive and negative samples. γ is the focusing parameter that adjusts the weighting scheme for samples, giving greater weight to hard-to-classify samples while lowering the weight of easy-to-classify ones. The pseudocode of ARDL in the training process is given in Figure 2.

Parameters: λ (redundancy parameter), bas (number of batches), eps (number of epochs).
Inputs: $\boldsymbol{R}$ (ratio groups), $\boldsymbol{E}$ (document embeddings), G (interfirm networks).

Initialize the main network and the conditional entropy estimators.
For epoch ← 1 to eps:
  for batch ← 1 to bas:
    if textual modality:
      Run a feedforward pass through a GRU based on $\boldsymbol{E}$ to get $\boldsymbol{H}$.
    Derive merged ratio group using the ratio-aware attention and feed it into an MLP layer to obtain financial representation $\boldsymbol{M}^f$.
    Derive textual representation $\boldsymbol{M}^t$ using the report-aware attention.
    Derive network representation $\boldsymbol{M}^n$ using the neighbor-aware attention.
    Compute $L_{F|T,N}$ with Eq. (16).
    Compute $L_{T|F,N}$ with Eq. (17).
    Compute $L_{N|F,T}$ with Eq. (18).
    Compute L_{fus} with Eq. (19).
    Concatenate $\boldsymbol{M}^f$, $\boldsymbol{M}^t$, and $\boldsymbol{M}^n$ and obtain $\hat{y}$ with Eq. (20).
    Minimize the total loss given in Eq. (21).

Figure 2. Pseudocode of ARDL.

4. Empirical Evaluation

4.1 Data Collection

To evaluate the effectiveness of our proposed method, we collected a multimodal dataset of companies in the NEEQ market of China from 2019 to 2021. The multimodal data of companies in 2019 was used to predict whether the companies getting into financial distress in 2021. Following previous studies (e.g., Wang et al., 2021; Zhang et al., 2022), we define "financial distress" as a company being given special treatment (ST). We excluded companies being ST in 2019 and 2020, resulting in 7,731 samples, including

7,366 normally operated companies and 365 companies being ST in 2021 (4.7% financial distress rate). For predictors of financial distress, we collected data for three modalities, including financial ratios, current reports, and interfirm networks.

We collected 30 financial ratios in the 2019 annual report of each company from the CSMAR (China Security Market Accounting Research) dataset. These financial ratios are widely used and proven to be effective for FDP in previous studies (Nazareth & Ramana Reddy, 2023). We divided the financial ratios into five groups based on the aspect of financial conditions they reflect, i.e., profitability, solvency, development capacity, operational capabilities, and finance structure. The descriptive statistics of these financial ratios are available in Appendix B.

For current reports, we collected all the current reports disclosed by each company from the official website of NEEQ[2] in 2019, resulting in 289,069 current reports. The mean and standard deviation of the number of current reports per company are 37.391 and 20.959, respectively. Considering that some companies used images instead of text in their current reports, we utilized Tesseract OCR technology to convert images to text. A detailed description and the distribution of current reports among companies are available in Appendix B.

For interfirm networks, we collected information regarding the neighbor companies of the 7,731 companies in our dataset from Qichacha[3] and built interfirm networks. Specifically, we first did a reverse lookup on the 7,731 companies to acquire information about their corresponding directors, supervisors, and senior management (DSS). We then collected the related companies of these DSS, resulting in 315,490 neighbor companies of the original 7,731 companies. A detailed description and the descriptive statistics of the DSS networks are available in Appendix B. For attributes to characterize each node (i.e., company) in DSS networks, we selected demographic attributes (i.e., company age, district, industry type, registered capital, number of patents, and percentage of insider ownership), following Long et al. (2022), and further

[2] NEEQ website (https://www.neeq.com.cn) offers publicly available information about NEEQ companies in China.
[3] Qichacha (https://www.qcc.com/) is a well-known enterprise information inquiry website in China.

considered three risk event attributes based on administrative penalties, equity pledges, and loan disputes, respectively. The risk event attributes were set to 1 if the company was involved in the corresponding risk events in 2019, and 0 otherwise.

4.2 Experimental Design

Financial distress prediction is generally treated as a binary classification task, i.e., identifying financially distressed companies from normal operation ones (Wang et al., 2021). We evaluated ARDL in comparison with benchmarked methods from two families, i.e., machine learning and deep learning. For machine learning, we used LR, SVM, XGB, and LightGBM. For deep learning, according to the way of using multimodal data, we used concatenation-based multimodal learning (CML) (Ngiam et al., 2011), low-rank multimodal fusion (LMF) (Liu et al., 2018), gated multimodal unit (GMU) (Du et al., 2022), Transformer-based multimodal network (TMN) (Korangi et al., 2023), attentive feature fusion (AFF) (Liu et al., 2022), and fine-grained attention network (FGAN) (Wang et al., 2023). Specifically, CML simply concatenates representations of all modalities together for prediction; LMF models multimodal interactions using tensor factorization; GMU adaptively filters out less important features across modalities using gate units; TMN learns intramodal interactions with the self-attention in Transformer encoder; AFF integrates multimodal representations using different weights given by the attention module; FGAN considers the intramodal and intermodal heterogeneities using fine-grained attentions.

In addition to predictive performance, we also compared ARDL with the six benchmarked deep learning methods in terms of representation performance (Chen et al., 2023). Specifically, we extracted the representation in the last layer of each deep learning method (including ARDL) as input features and built four types of machine learning methods (i.e., LR, SVM, XGB, and LightGBM). The predictive performance of these machine learning methods could therefore reflect the quality of the representation generated by each deep learning method.

To determine the optimal hyperparameters for both ARDL and benchmarked methods, we employed nested cross-validation (10-fold split in both inner and outer loops) with grid search for parameter tuning. Specifically, we first divided the original dataset into ten folds, with one fold for testing and the remaining

nine folds for training. Then, we further divided the training set into ten folds, with one fold as a validation set and the remaining nine folds as a reduced training set. Such process was repeated ten times in both inner and outer loops. We used average performance (in terms of AUC) over validation sets to select the hyperparameters for each method. We performed ten independent nested cross-validations to get a robust result. Method implementation and parameter settings are available in Appendix C.

We selected five metrics for comprehensively gauging the predictive performance of each method: the area under the receiver operating characteristic curve (AUC), Kolmogorov-Smirnov (KS) statistic, H measure, precision, and recall (Hand, 2009). To estimate out-of-sample performance of each method, we performed ten independent 10-fold cross-validations, resulting in 100 performance estimates, to get a robust result. The performance results (mean and standard deviation) reported later are all based on the 100 estimates. For a fair comparison between methods, the partitioning of folds was kept identical across all methods during each 10-fold cross-validation.

We also performed a series of additional experiments to analyze the performance of ARDL on all fronts. Specifically, an ablation study was performed to examine whether and how each design artifact influences the predictive performance; an interpretation analysis was performed to demonstrate how ARDL adaptively identifies important information using attention weights; a sensitivity analysis was performed to reveal how the conditional entropy-based regularization impacts ARDL's performance; two robustness tests were conducted to examine whether the utility of ARDL is affected by dataset selection and data diversity.

5. Experimental Results and Analysis

5.1 Predictive Performance

We first examined the predictive performance of each method using different degrees of multimodal data (i.e., single modality, dual modalities, and triple modalities). For current reports, we obtained a 768-dimensional document vector of each report using BERT. We calculated the mean of all document vectors of each company as the input of textual modality for machine learning methods and deep learning methods without temporal modeling (AFF, CML, LMF, and GMU). We used a sequence of document vectors of each company as the input of textual modality for deep learning methods with temporal modeling (TMN,

FGAN, and ARDL). For interfirm networks, we used the graph neural network to generate a 128-dimensional feature vector (i.e., node embedding) for each company as the input of network modality for both machine learning and deep learning methods. Table 3 summarizes the results of each method (mean and standard deviation) using single modality, dual modalities, and triple modalities, respectively. Additionally, considering such a high dimension of textual input may be intractable for a machine learning method, we reduced the dimension of the feature vector to 50 using principal component analysis (PCA) (Jiang et al., 2022a), and used the reduced feature vector as the input in terms of textual modality for the machine learning methods. The detailed results of each method are available in Appendix D. Meanwhile, considering the sentiment and readability information embedded in current reports may provide additional signals for the textual input, we further extracted two types of sentiment and readability features, respectively. These features were then combined with the semantic features extracted by BERT to construct the textual input for both ARDL and benchmarks. The detailed descriptions of the extraction process and the results of each method are available in Appendix D.

The results show that ARDL significantly outperformed all benchmarked methods, under unimodal, bimodal, and trimodal inputs. This demonstrates the advantages of our method, which is designed by considering the intra-modality importance and the inter-modality heterogeneity, as well as the complementarity and redundancy of modality information for FDP. To further clarify the experimental results, we make vertical and horizontal comparative analyses.

From the vertical perspective (comparison among modalities), we find that incorporating the textual or network features significantly improved performance for every method compared to relying solely on financial features. Furthermore, the use of a combination of financial and network features always outperformed the combination of financial and textual features, indicating that network features based on relational risks and demographic data could be more valuable than textual features based on disclosure reports data in FDP. By incorporating trimodal features, the models always achieved the best performance, showing the effectiveness of using trimodal data. An interesting finding lies in that for some methods (e.g., LR and SVM), trimodal input not always yielded better performance than bimodal inputs. The potential

Table 3. Predictive Performance of ARDL versus Benchmarks (%).

Method	Fin					Fin+Text					Fin+Net					Fin+Text+Net				
Metrics	AUC	KS	H	Precise	Recall	AUC	KS	H	Precise	Recall	AUC	KS	H	Precise	Recall	AUC	KS	H	Precise	Recall
LR	79.10	58.36	45.03	77.89	81.62	83.97	62.13	47.02	79.54	81.66	85.50	64.43	50.64	80.35	83.72	81.96	61.63	46.63	78.31	81.07
	(1.91)	(1.74)	(1.86)	(1.74)	(1.84)	(2.73)	(2.76)	(2.93)	(2.81)	(2.84)	(2.72)	(2.84)	(2.59)	(2.47)	(2.64)	(2.34)	(2.17)	(2.35)	(2.29)	(2.36)
SVM	82.28	58.68	45.33	80.19	81.37	84.63	63.49	49.50	81.27	83.48	85.84	65.65	51.59	82.38	85.40	83.63	63.80	52.62	80.14	83.73
	(1.85)	(1.83)	(1.57)	(1.74)	(1.76)	(2.26)	(2.64)	(2.51)	(2.27)	(2.45)	(1.93)	(2.43)	(2.71)	(1.83)	(1.98)	(2.36)	(2.61)	(2.47)	(1.79)	(1.78)
XGB	84.68	61.09	46.88	82.82	84.04	86.21	64.94	50.05	82.00	83.85	87.14	66.65	53.00	82.87	85.79	87.67	68.55	54.63	81.68	84.43
	(1.50)	(1.53)	(1.76)	(2.12)	(2.12)	(2.31)	(2.18)	(2.47)	(2.35)	(2.87)	(1.98)	(2.15)	(2.21)	(2.20)	(1.96)	(1.58)	(2.16)	(1.84)	(2.10)	(2.18)
LightGBM	83.96	59.72	46.56	81.90	83.92	85.63	64.22	51.41	82.93	85.03	86.99	64.99	52.76	83.70	86.69	87.49	67.68	54.38	82.69	85.35
	(1.43)	(1.78)	(1.69)	(1.96)	(1.97)	(2.03)	(2.73)	(2.42)	(2.18)	(2.74)	(2.12)	(1.98)	(2.65)	(2.04)	(2.12)	(2.04)	(2.03)	(1.81)	(2.08)	(1.93)
TMN	85.34	64.17	52.07	82.15	85.30	87.58	65.57	51.98	83.51	82.08	88.72	67.60	53.62	84.44	86.03	89.77	69.45	56.72	83.19	85.65
	(1.12)	(1.21)	(1.46)	(2.01)	(2.02)	(1.60)	(1.46)	(1.84)	(2.34)	(2.83)	(1.20)	(1.70)	(1.63)	(1.96)	(2.17)	(1.36)	(1.60)	(1.41)	(1.88)	(1.91)
AFF	85.01	62.12	49.31	82.32	85.05	87.05	65.40	54.05	84.31	86.24	87.32	67.43	53.49	85.19	86.92	88.76	69.75	55.79	84.35	86.39
	(1.04)	(1.48)	(1.10)	(1.67)	(1.59)	(1.77)	(1.51)	(1.75)	(2.08)	(1.93)	(1.38)	(1.41)	(1.27)	(1.70)	(1.79)	(1.76)	(1.92)	(1.70)	(1.58)	(1.93)
FGAN	85.18	63.30	50.05	83.20	85.50	88.29	66.81	54.40	85.16	86.80	88.81	68.81	56.40	85.78	87.14	90.94	71.13	59.90	84.86	85.99
	(1.18)	(0.95)	(1.17)	(1.59)	(1.86)	(1.39)	(1.73)	(2.18)	(2.06)	(2.27)	(1.74)	(1.83)	(1.61)	(1.69)	(1.62)	(1.63)	(1.89)	(1.83)	(1.74)	(1.84)
CML	-	-	-	-	-	86.94	65.21	50.90	84.08	85.49	87.67	66.76	55.55	84.53	87.43	89.27	69.83	56.75	83.71	85.88
						(1.62)	(1.78)	(2.03)	(2.37)	(2.01)	(1.66)	(1.85)	(1.52)	(2.31)	(2.22)	(1.33)	(2.20)	(1.47)	(2.05)	(2.20)
LMF	-	-	-	-	-	86.80	66.60	52.93	83.56	85.45	86.94	65.09	52.66	84.20	86.85	88.78	68.49	55.83	81.20	83.59
						(1.78)	(2.09)	(1.68)	(2.11)	(2.21)	(1.57)	(1.91)	(1.65)	(1.99)	(1.87)	(1.91)	(2.11)	(1.55)	(1.95)	(1.96)
GMU	-	-	-	-	-	87.63	67.07	53.54	84.72	86.57	88.67	67.64	55.99	85.41	87.06	90.20	70.02	58.32	83.54	86.76
						(1.51)	(1.76)	(1.50)	(2.06)	(1.88)	(1.08)	(1.70)	(1.42)	(1.69)	(1.91)	(1.66)	(2.21)	(1.68)	(1.69)	(1.78)
ARDL	**85.84**	**65.24**	**52.35**	**84.50**	**85.88**	**89.25**	**68.27**	**56.10**	**85.54**	**87.16**	**90.25**	**69.30**	**57.45**	**86.35**	**88.51**	**91.76**	**72.58**	**62.19**	**85.37**	**87.39**
	(0.80)	(0.85)	(0.60)	(1.16)	(1.01)	(1.47)	(1.66)	(1.27)	(1.63)	(1.84)	(1.12)	(1.52)	(1.24)	(1.29)	(1.68)	(1.48)	(1.58)	(1.62)	(2.11)	(1.96)

Notes: “Fin” refers to financial indicators; “Text” refers to current reports; “Net” refers to interfirm networks. The best performance is in boldface.

Table 4. Representation Performance of ARDL versus Benchmarks (%).

Method	LR					SVM					XGB					LightGBM				
Metrics	AUC	KS	H	Precise	Recall	AUC	KS	H	Precise	Recall	AUC	KS	H	Precise	Recall	AUC	KS	H	Precise	Recall
TMN	90.57	68.90	59.00	83.44	84.63	90.58	68.73	57.49	83.25	83.83	91.82	70.61	58.33	83.78	84.76	90.82	70.86	59.20	84.07	85.36
	(1.45)	(1.49)	(1.37)	(1.22)	(1.50)	(1.52)	(1.56)	(1.43)	(1.41)	(1.45)	(1.45)	(1.40)	(1.15)	(1.23)	(1.38)	(1.32)	(1.36)	(1.10)	(1.34)	(1.19)
AFF	89.52	68.43	58.27	83.27	83.93	89.63	68.34	56.35	82.59	83.58	91.21	71.07	59.99	84.38	83.96	90.44	70.87	59.45	84.78	84.60
	(1.31)	(1.65)	(1.60)	(1.41)	(1.48)	(1.42)	(1.77)	(1.58)	(1.70)	(1.69)	(1.36)	(1.59)	(1.44)	(1.46)	(1.36)	(1.34)	(1.61)	(1.46)	(1.26)	(1.30)
FGAN	90.90	70.61	60.63	84.65	85.39	90.68	70.66	60.51	84.10	84.81	92.00	71.91	62.74	85.28	85.10	91.67	72.37	62.88	86.27	86.74
	(1.34)	(1.68)	(1.32)	(1.13)	(1.07)	(1.41)	(1.93)	(1.27)	(1.30)	(1.21)	(1.25)	(1.92)	(1.05)	(1.00)	(1.22)	(1.19)	(1.87)	(1.19)	(1.11)	(1.09)
CML	89.31	67.64	56.44	82.55	83.68	88.58	68.12	55.59	82.36	82.97	89.95	69.32	56.20	84.21	83.68	90.08	69.99	56.91	84.27	84.59
	(1.61)	(2.07)	(1.63)	(1.51)	(1.37)	(1.74)	(2.02)	(1.85)	(1.70)	(1.75)	(1.82)	(2.08)	(1.77)	(1.67)	(1.40)	(1.53)	(1.96)	(1.73)	(1.52)	(1.45)
LMF	88.88	66.15	55.51	82.58	82.15	88.49	65.54	55.57	81.98	82.92	89.62	66.98	56.38	82.90	82.55	89.65	67.88	56.56	83.66	83.89
	(1.50)	(2.08)	(1.53)	(1.61)	(1.69)	(1.71)	(1.95)	(1.81)	(1.79)	(1.65)	(1.64)	(1.78)	(1.72)	(1.68)	(1.76)	(1.83)	(1.92)	(1.47)	(1.74)	(1.52)
GMU	90.40	70.57	60.07	84.14	84.88	89.27	70.90	60.92	83.44	84.14	90.89	71.67	62.75	84.30	85.12	90.40	71.93	61.78	84.44	86.44
	(1.53)	(2.88)	(1.34)	(1.19)	(1.35)	(1.74)	(2.12)	(1.43)	(1.44)	(1.33)	(1.55)	(2.05)	(1.22)	(1.00)	(1.15)	(1.39)	(2.18)	(1.23)	(1.04)	(1.24)
ARDL	**91.43**	**71.39**	**61.63**	**85.29**	**86.39**	**91.84**	**71.72**	**62.71**	**84.88**	**85.15**	**92.87**	**74.03**	**64.69**	**86.30**	**88.17**	**92.32**	**72.44**	**63.91**	**86.70**	**88.26**
	(1.27)	(1.69)	(0.95)	(1.03)	(0.95)	(1.42)	(1.89)	(1.18)	(1.18)	(1.18)	(1.12)	(1.85)	(1.28)	(1.03)	(1.13)	(1.23)	(1.92)	(1.25)	(0.99)	(1.00)

Note: The best performance is in boldface.

reason may be that neglecting intramodal heterogeneity and simply concatenating multimodal inputs led to overfitting. Such finding further highlights the importance of extracting key information within and among modalities in FDP.

From the horizontal perspective (comparison among methods), we find that, compared to machine learning methods, deep learning methods always made noticeable improvements on every data set. Among the deep learning benchmarks, FGAN, which incorporates intramodal and intermodal attentions, outperformed others for every data input, manifesting the advantage of attention mechanisms for dealing with heterogeneous utilities implied in multimodal data. On the four data sets, ARDL yielded improved performance over all attention-based benchmarks (i.e., TMN, AFF, and FGAN). This result demonstrates the facilitating role of the modality-specific attentions and the conditional entropy-based regularization on prediction performance. ARDL also outperformed gate-based GMU, concatenation-based CML, and factorization-based LMF. Finally, ARDL consistently surpassed all the machine learning benchmarks, including linear and ensemble methods.

5.2 Representation Performance

Given the capability of representation learning of multimodal data, we further examined the representation performance of ARDL versus benchmarked deep learning methods. Specifically, we extracted the representation of the last layer of each method and built four downstream classifiers (LR, SVM, XGB, and LightGBM) for FDP. Hence, the better the representation performance, the better the predictive performance of the downstream classifiers. Table 4 summarizes the results of representation performance of each method using trimodal data.

From the vertical perspective (comparison among classifiers), ensemble methods (XGB and LightGBM) achieved better performance than linear methods (LR and SVM) across all representations and metrics. Notably, XGB led to the best performance, demonstrating its superior classification capability for multimodal-based FDP. From the horizontal perspective (comparison among representations), the results show that the predictive performance of the downstream classifiers of ARDL was significantly superior to those of all benchmarked methods, indicating a better representation performance. As with the results of

the predictive performance, representations generated by FGAN achieved the best performance among the deep learning benchmark methods. Another interesting finding lies in that with the representations learned by the deep learning methods, the downstream classifiers (XGB and LightGBM) may yield even better predictive performance. For example, the mean AUC of XGB using the representation of ARDL was 92.87, higher than the original ARDL (91.76). The reason may be that ensemble learning works better for the final tabular classification task, as compared to MLP (Gorishniy et al., 2022). Such finding also indicates a promising direction, i.e., designing two-stage methods for FDP, consisting of a deep learning module for generating joint modality representation and an ensemble learning module for the downstream prediction task.

5.3 Ablation Study

To examine whether and how each design artifact affects the performance of ARDL, we performed an ablation study. Specifically, we built six ablative variants (M_1 to M_6) of ARDL, each of which drops a design artifact in ARDL. Methods M_1, M_2, and M_3 drop the ratio-aware, report-aware, and neighbor-aware attention modules, respectively. Method M_4 drops all the three modality-specific attention modules. Method M_5 drops the conditional entropy-based regularized fusion module. Method M_6 drops the focal loss module and uses the cross-entropy loss instead. Figure 3 illustrates the results of the six ablative variants in terms of AUC (mean values across ten times 10-fold cross-validations).

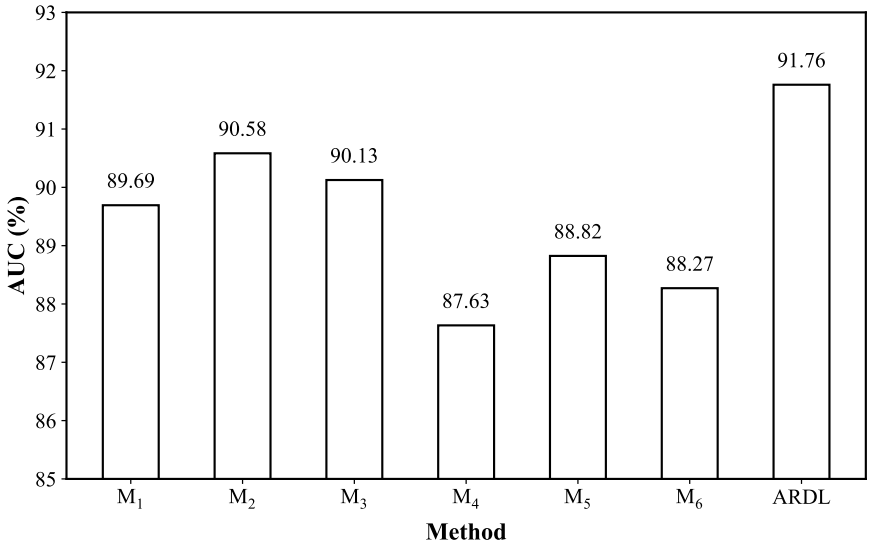

Figure 3. Results of Ablation Study.

The results show that M_1, M_2, and M_3 exhibit varying degrees of performance degradation in terms of AUC as compared to ARDL and there is a significant gap between M_4 and ARDL, indicating that each modality-specific attention module contributes to performance improvement of FDP. Besides, dropping either conditional entropy-based regularization (M_5) or focal loss (M_6) both show significant performance degradation. In summary, the results of ablation study provide clear evidence that all the design artifacts (modality-specific attentions, conditional entropy-based regularization, and focal loss) contribute to performance improvement and they collectively ensure the effectiveness of ARDL.

5.4 Interpretation Analysis on Attention Weights

To examine whether and how each modality-specific attention identifies key information, we also performed an interpretation analysis on attention weights. The attention weights show the distinct importance of different financial ratio groups, current reports, and neighbor companies and help to better understand the factors contributing to FDP. Figures 4 to 6 illustrate the average attention weights of financial ratio groups, current reports, and neighbor companies, respectively, across ten independent 10-fold cross-validations (each block presents the average attention weight across one run of 10-fold cross-validations).

For financial ratios, we divided them into five groups based on the financial aspects they reflect (i.e., profitability, solvency, development capacity, operational capabilities, and finance structure). For example, profitability reflects a company's ability to generate profits, including metrics like return on assets and return on equity. Solvency reflects a company's ability to meet its financial obligations, using measures such as the current ratio and quick ratio. Both of these two groups play important roles in distinguishing between normally operated and financially distressed companies, as compared to other financial ratio groups such as development capacity and operational capabilities. Such finding is consistent with prior studies (Fu et al., 2020), in which financial distress of a company has been found to be generally associated with low profitability and solvency. In addition, the financial structure ratio group, which reflects a company's capital and debt arrangement, may be less indicative of financial distress as some of the information it provides may already be captured by other indicators. This is demonstrated by its attention weights consistently being at a low level.

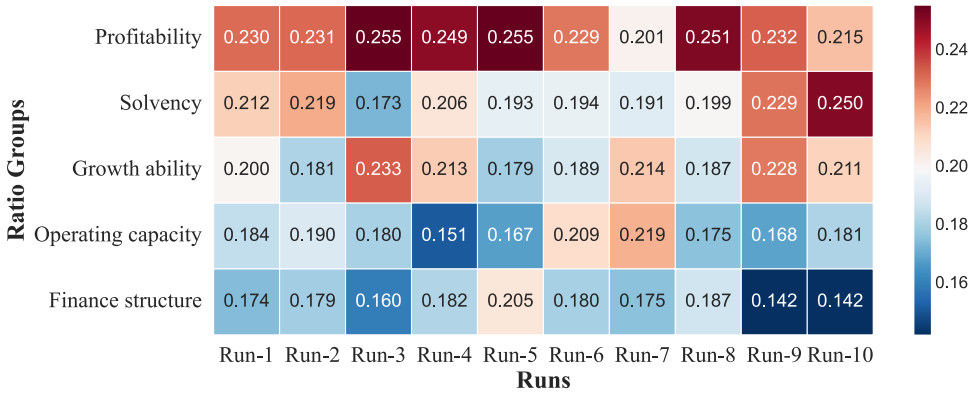

Figure 4. Attention Weights of Financial Ratio Groups.

For current reports, we divided them into ten categories based on their specific disclosure scenarios. It should be noted that the division of current reports is only intended to facilitate the interpretive analysis. Among the categories, *corporate operations* refer to the current reports related to business operations, such as outward investments, provision of guarantees, and interfirm transactions. *Litigation and lawsuit* refer to the current reports related to material risk events that companies are involved, including legal proceedings, arbitrations, and violations. We then calculated the average attention weights for each category of report (illustrated in Figure 5). The results show that both *corporate operations* and *litigation and lawsuit* related reports show outstanding predictive ability. The reason may lie in that corporate operations-related reports provide valuable information reflecting financial position and overall stability of the company. Litigation and lawsuit-related reports reflect the risk events that could adversely affect the financial position of the company (Yin et al., 2020). Besides, reports on other events, such as stock issuance and notice of general meetings, were found to be less indicative of financial distress, as they offer limited insights into the financial position of the company.

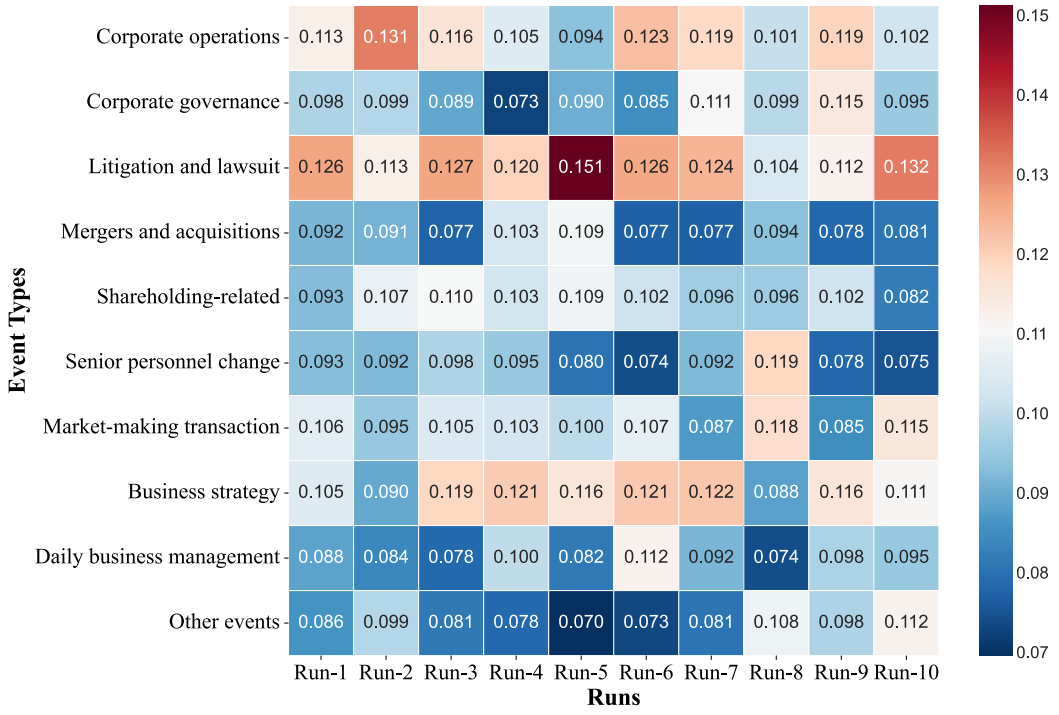

Figure 5. Attention Weights of Current Report Groups.

For interfirm networks, we divided neighbor companies into four groups based on the types of risk events (i.e., administrative penalties, equity pledges, and loan disputes) they were involved in. It should be noted that the division of neighbors is only intended to facilitate the interpretive analysis. Among the groups, neighbor companies-3 refers to the neighbor companies that were involved in all three types of risk events within a year. Neighbor companies-2 refers to the neighbor companies that were involved in any two types of risk events within a year, such as administrative penalties and equity pledges, or equity pledges and loan disputes. We then calculated the average attention weights for each group (illustrated in Figure 6). The results show a positive correlation between the risk levels of neighbor companies and the assigned attention weights, indicating that companies with higher-risk neighbor companies are more prone to financial distress due to the contagion effect of risk events. The neighbor companies-3 exhibit a remarkable capability in predicting the financial distress of target companies, as evidenced by their consistently superior attention weights across all runs. They are followed by neighbor companies-2 and neighbor companies-1 in terms of predictive power. In contrast, the neighbor companies-0, which have no risk events, contribute less to FDP compared to other companies. In this regard, the elevated weights assigned to riskier neighbor companies

provide valuable guidance for stakeholders to assess the level of attention they should allocate to specific counterparts, and thus help them identify more valuable financial distress clues.

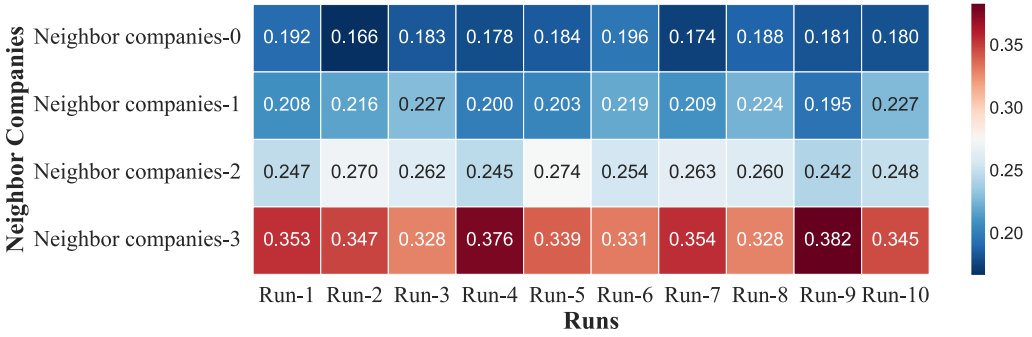

Figure 6. Attention Weights of Neighbor Company Groups.

5.5 Sensitivity Analysis

To investigate the effect of the regularization parameter λ on the performance of ARDL, we also performed a sensitivity analysis. A high value of λ may lead to inadequate representation of multimodal information, whereas a low value of λ may lead to an inadequate focus on redundant information. We set the value of λ with different values $(1, 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4})$ and examined the performance of ARDL in terms of AUC (illustrated in Figure 7). The results show that the performance of ARDL gradually improved with the value of λ decreased from 1 to 10^{-2} and then the performance reduced with the value of λ lower than 10^{-2}. The empirical results suggest that 10^{-2} may be a reasonable value to yield an effective tradeoff between the two losses. We caution that the optimal regularization parameter may vary across contexts and finding an appropriate regularization parameter may be necessary when applying ARDL for other contexts. If the modeling scenario undergoes significant change, the parameter λ remains amenable to adjustment through parameter tuning.

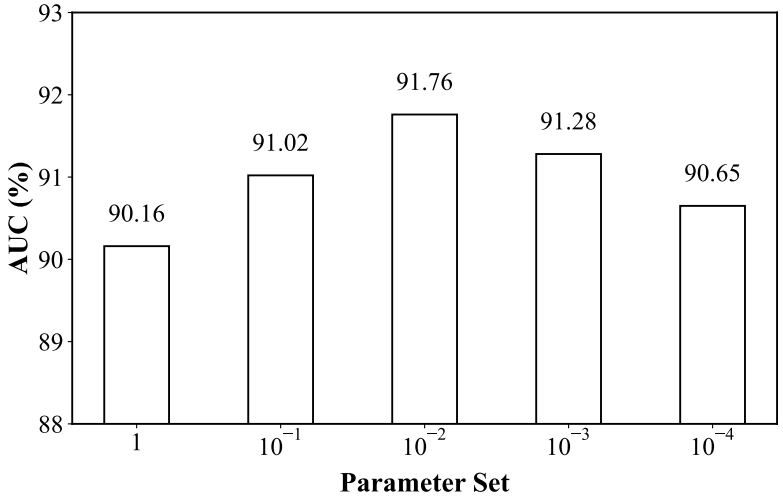

Figure 7. Results of Sensitivity Analysis.

5.6 Robustness Tests

We conducted two robustness tests to further examine whether the utility of our proposed method is affected by (1) dataset selection; (2) data diversity. Overall, the results demonstrate that the effectiveness of ARDL is robust across different contexts and higher diversity of data sources.

The first test examines whether the effectiveness of ARDL is robust to the potential dataset selection bias. Specifically, we first collected an additional multimodal dataset of listed companies in the Shenzhen Stock Exchange and Shanghai Stock Exchange of China from 2019 to 2022 for FDP. A detailed description of the multimodal dataset is available in Appendix E. We then examined the robustness of ARDL on the new dataset by comparing its predictive and representation performance with benchmarked methods. The detailed results of each method are available in Appendix F. Overall, the results show similar patterns in the two multimodal cases (i.e., NEEQ companies and listed companies). ARDL consistently outperformed all benchmarked methods in terms of predictive and representation performance in both multimodal cases. This demonstrates that the effectiveness and generalization of our method under different contexts. Additionally, incorporating multimodal features improved performance for every method, showing that the performance improvement due to the use of multimodal features was also robust.

The second test examines whether the effectiveness of ARDL is robust with higher diversity of data sources. First of all, considering the three attention modules of ARDL are designed for the specific modality

representations in multimodal contexts, they may not be suitable for multi-view datasets where the data structures do not necessarily conform to those modality assumptions. We thereby built a general variant of ARDL (ARDLv) by dropping the three modality-specific attentions for FDP. To have a fair comparison, we also removed the data representation modules of the benchmarked methods (TMN, FGAN) and built their variants (TMNv, FGANv) for FDP. Besides, we further collected a more diverse multi-view dataset of listed companies from five distinct sources in the Shenzhen Stock Exchange and Shanghai Stock Exchange of China from 2019 to 2022 for FDP. A detailed description of the multi-view dataset is available in Appendix G. Then, we examined the robustness of the general variant of ARDL (i.e., combining the conditional entropy-based regularized fusion module and the focal loss module) on the multi-view dataset by comparing its predictive and representation performance with benchmarked methods. The detailed results of each method are available in Appendix H. Overall, ARDL outperformed all benchmarked methods in terms of predictive and representation performance, validating the usefulness and stability of our proposed method on more diverse data sources.

5.7 Findings and Implications

Our work contributes significant findings from the aspects of information and methodology. In regard to information for financial distress prediction, our study demonstrates that the three identified modalities indeed have discriminative, albeit heterogeneous, ability in predicting financial distress. These modalities ranked from high to low in terms of predictive performance are financial indicators, interfirm networks, and current reports. Financial indicators, particularly profitability and solvency-related ratios, are highly indicative of financial distress and provide valuable insights into potential issues related to debt repayment. Echoing the evidence that the failure or distress of one company may trigger a domino effect due to interrelation and interdependencies within the financial system, our study clearly demonstrates that companies with riskier neighbors are more susceptible to financial distress due to the contagion effect of risk events. Additionally, among various types of current reports, corporate operations and litigation and lawsuit-related reports are found to be the most informative in reducing information asymmetry between companies and stakeholders, thereby mitigating financial distress.

From the aspect of methodology, on the one hand, our study highlights the superior predictive performance of leveraging multimodal data for FDP; on the other hand, leveraging all three modalities does not always result in superior performance compared to using only two modalities, especially for linear methods (e.g., LR and SVM). One potential explanation lies in that neglecting the heterogeneity within each modality and simply concatenating multimodal inputs may impede the extraction of valuable knowledge from multimodal data. This finding further emphasizes the importance of identifying and extracting key information within and among modalities in FDP. Additionally, we found that the two-stage method may even yield superior predictive performance, i.e., a deep learning module for generating a joint modality representation and then an ensemble learning module for predicting financial distress.

Our work also provides practical implications for stakeholders. First, regarding the benefits of utilizing multimodal data for FDP (e.g., financial indicators, current reports, and interfirm networks), stakeholders (e.g., investors) may consider extracting valuable information from multiple sources to gain a comprehensive understanding of a company's risk profile. Second, the utilization of flexible modeling methods (e.g., ARDL) that treat each modality as an independent source of knowledge and promote the integration of complementary information from multiple modalities during fusion, is essential when predicting financial distress using multimodal data. Stakeholders (e.g., financial institutions) can leverage the proposed method to enhance the performance of FDP, thus leading to improved investment decisions and loan assessments. Moreover, companies can enhance investor and creditor confidence by disclosing additional information, such as ESG reports, to mitigate information asymmetry.

6. Conclusion

In light of the increasing value of multimodal data and the challenges of leveraging multimodal data for FDP, we propose an attentive and regularized deep learning method. The proposed method synthesizes three designed modality-specific attention modules to explicitly and adaptively extract key information within each modality and a novel conditional entropy-based regularization to alleviate redundant information during modality fusion. Our empirical evaluation results demonstrate the advantage of ARDL. It significantly outperformed benchmarked methods in terms of both predictive performance and

representation performance. Ablation study confirms the performance improvement effects of its core components. Interpretation analysis on attention weights illustrates how each feature within each modality differentially worked for FDP. Sensitivity analysis reveals how the regularization parameter effect the performance of ARDL. Robustness tests clearly show that the utility of ARDL is not affected by dataset selection and data diversity. We also provide the key findings and implications.

Our study has several limitations that could be addressed in future research. First, we used three representative modalities to predict financial distress and there exist numerous other informative modalities, such as earnings conference calls (Li et al., 2020) and initial public offering roadshow videos (Freiberg and Matz, 2023). Future research may explore the utility of new data modalities and expand ARDL to leverage additional modalities, for better predictive performance. Second, we treated financial distress prediction as a binary classification problem, whereas financial distress could be more nuanced, ranging from slight financial distress to severe financial distress (e.g., bankruptcy). Future research may consider extending ARDL to accommodate various degrees of financial distress.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (grants 71731005 and 72101073) and the Anhui Provincial Natural Science Foundation (grant 2108085MG234).

Reference

Beaver, W. H., Cascino, S., Correia, M., & McNichols, M. F. (2019). Group affiliation and default prediction. *Management Science*, *65*(8), 3559–3584.

Bi, W., Xu, B., Sun, X., Wang, Z., Shen, H., & Cheng, X. (2022). Company-as-tribe: Company financial risk assessment on tribe-style graph with hierarchical graph neural networks. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining* (pp. 2712–2720).

Borchert, P., Coussement, K., De Caigny, A., & De Weerdt, J. (2023). Extending business failure prediction models with textual website content using deep learning. *European Journal of Operational Research*, *306*(1), 348–357.

Chen, G., Xiao, S., Zhang, C., & Zhao, H. (2023). A theory-driven deep learning method for voice chat–based customer response prediction. *Information Systems Research*, isre.2022.1196.

Chen, N., Ribeiro, B., & Chen, A. (2016). Financial credit risk assessment: A recent review. *Artificial Intelligence Review*, *45*(1), 1–23.

Cui, G., & Li, Y. (2022). Nonredundancy regularization based nonnegative matrix factorization with manifold learning for multiview data representation. *Information Fusion*, *82*, 86–98.

Dastile, X., Celik, T., & Potsane, M. (2020). Statistical and machine learning models in credit scoring: A systematic literature survey. *Applied Soft Computing*, *91*, 106263.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Dikmen, M., & Burns, C. (2022). The effects of domain knowledge on trust in explainable AI and task performance: A case of peer-to-peer lending. *International Journal of Human-Computer Studies*, *162*, 102792.

Du, Y., Liu, Y., Peng, Z., & Jin, X. (2022). Gated attention fusion network for multimodal sentiment classification. *Knowledge-Based Systems*, *240*, 108107.

Fan, S., & Ilk, N. (2020). A text analytics framework for automated communication pattern analysis. *Information & Management*, *57*(4), 103219.

Freiberg, B., & Matz, S. C. (2023). Founder personality and entrepreneurial outcomes: A large-scale field study of technology startups. *Proceedings of the National Academy of Sciences*, *120*(19), e2215829120.

Fu, X., Ouyang, T., Chen, J., & Luo, X. (2020). Listening to the investors: A novel framework for online lending default prediction using deep learning neural networks. *Information Processing & Management*, *57*(4), 102236.

Ge, P., Ren, C.-X., Xu, X.-L., & Yan, H. (2023). Unsupervised domain adaptation via deep conditional adaptation network. *Pattern Recognition*, *134*, 109088.

Geng, R., Bose, I., & Chen, X. (2015). Prediction of financial distress: An empirical study of listed Chinese companies using data mining. *European Journal of Operational Research*, *241*(1), 236–247.

González-López, J., Ventura, S., & Cano, A. (2020). Distributed selection of continuous features in multilabel classification using mutual information. *IEEE Transactions on Neural Networks and Learning Systems*, *31*(7), 2280–2293.

Gorishniy, Y., Rubachev, I., & Babenko, A. (2022). On embeddings for numerical features in tabular deep learning. *Advances in Neural Information Processing Systems*, *35*, 24991–25004.

Hand, D. J. (2009). Measuring classifier performance: A coherent alternative to the area under the ROC curve. *Machine Learning*, *77*(1), 103–123.

Huang, Y., Du, C., Xue, Z., Chen, X., Zhao, H., & Huang, L. (2021). What makes multi-modal learning better than single (provably). *Advances in Neural Information Processing Systems*, *34*, 10944–10956.

Iyer, R., Khwaja, A. I., Luttmer, E. F. P., & Shue, K. (2016). Screening peers softly: Inferring the quality of small borrowers. *Management Science*, *62*(6), 1554–1577.

Jiang, C., Lyu, X., Yuan, Y., Wang, Z., & Ding, Y. (2022a). Mining semantic features in current reports for financial distress prediction: Empirical evidence from unlisted public firms in China. *International Journal of Forecasting*, *38*(3), 1086–1099.

Jiang, C., Wang, J., Wang, Z., Liu, X. (2022b). Capturing heterogeneous interactions for financial risk prediction of SMEs. *PACIS 2022 Proceedings*, *56*, 1361.

Korangi, K., Mues, C., & Bravo, C. (2023). A transformer-based model for default prediction in mid-cap corporate markets. *European Journal of Operational Research*, *308*(1), 306–320.

Kou, G., Xu, Y., Peng, Y., Shen, F., Chen, Y., Chang, K., & Kou, S. (2021). Bankruptcy prediction for SMEs using transactional data and two-stage multiobjective feature selection. *Decision Support Systems*, *140*, 113429.

Kraus, M., & Feuerriegel, S. (2017). Decision support from financial disclosures with deep neural networks and transfer learning. *Decision Support Systems*, *104*, 38–48.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444.

Lee, J. W., Lee, W. K., & Sohn, S. Y. (2021). Graph convolutional network-based credit default prediction utilizing three types of virtual distances among borrowers. *Expert Systems with Applications*, *168*, 114411.

Li, J., Yang, L., Smyth, B., & Dong, R. (2020). MAEC: A multimodal aligned earnings conference call dataset for financial risk prediction. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management* (pp. 3063–3070).

Li, S., Shi, W., Wang, J., & Zhou, H. (2021). A deep learning-based approach to constructing a domain sentiment lexicon: A case study in financial distress prediction. *Information Processing & Management*, *58*(5), 102673.

Lin, T.-Y., Goyal, P., Girshick, R., He, K., & Dollar, P. (2020). Focal loss for dense object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *42*(2), 318–327.

Liu, J., Wang, S., Ma, W.-C., Shah, M., Hu, R., Dhawan, P., & Urtasun, R. (2020). Conditional entropy coding for efficient video compression. In *European Conference on Computer Vision* (pp. 453-468).

Liu, X., Li, Y., Jiang, C., Wang, Z., Zhao, F., & Wang, J. (2022). Attentive feature fusion for credit default prediction. In *2022 IEEE 25th International Conference on Computer Supported Cooperative Work in Design* (pp. 816–821).

Liu, Z., Shen, Y., Lakshminarasimhan, V. B., Liang, P. P., Zadeh, A., & Morency, L.-P. (2018). Efficient low-rank multimodal fusion with modality-specific factors. *arXiv preprint arXiv:1806.00064*.

Long, J., Jiang, C., Dimitrov, S., & Wang, Z. (2022). Clues from networks: Quantifying relational risk for credit risk evaluation of SMEs. *Financial Innovation*, *8*(1), 1–41.

Lu, C., Zhou, G., & Li, M. (2023). Research on information fusion method for heat model and weather model based on HOGA-SVM. *Multimedia Tools and Applications*, *82*(6), 9381–9398.

Matin, R., Hansen, C., Hansen, C., & Mølgaard, P. (2019). Predicting distresses using deep learning of text segments in annual reports. *Expert Systems with Applications*, *132*, 199–208.

Medina-Olivares, V., Calabrese, R., Dong, Y., & Shi, B. (2022). Spatial dependence in microfinance credit default. *International Journal of Forecasting*, *38*(3), 1071–1085.

Nazareth, N., & Ramana Reddy, Y. V. (2023). Financial applications of machine learning: A literature review. *Expert Systems with Applications*, *219*, 119640.

Ngiam, J., Khosla, A., Kim, M., Nam, J., Lee, H., & Ng, A. Y. (2011). Multimodal deep learning. In *Proceedings of the 28th International Conference on International Conference on Machine Learning* (pp. 689–696).

Pichler, G., Colombo, P. J. A., Boudiaf, M., Koliander, G., & Piantanida, P. (2022). A differential entropy estimator for training neural networks. In *Proceedings of the 39th International Conference on Machine Learning* (pp. 17691–17715).

Qian, H., Wang, B., Yuan, M., Gao, S., & Song, Y. (2022). Financial distress prediction using a corrected feature selection measure and gradient boosted decision tree. *Expert Systems with Applications*, *190*, 116202.

Schmid, L., Gerharz, A., Groll, A., & Pauly, M. (2023). Tree-based ensembles for multi-output regression: Comparing multivariate approaches with separate univariate ones. *Computational Statistics & Data Analysis*, *179*, 107628.

Shalev, Y., Painsky, A., & Ben-Gal, I. (2022). Neural joint entropy estimation. *IEEE Transactions on Neural Networks and Learning Systems*, 1–13.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*(3), 379–423.

Sun, J., Fujita, H., Zheng, Y., & Ai, W. (2021). Multi-class financial distress prediction based on support vector machines integrated with the decomposition and fusion methods. *Information Sciences*, *559*, 153–170.

Sun, J., Li, H., Fujita, H., Fu, B., & Ai, W. (2020). Class-imbalanced dynamic financial distress prediction based on Adaboost-SVM ensemble combined with SMOTE and time weighting. *Information Fusion*, *54*, 128–144.

Tobback, E., Bellotti, T., Moeyersoms, J., Stankova, M., & Martens, D. (2017). Bankruptcy prediction for SMEs using relational data. *Decision Support Systems*, *102*, 69–81.

Topuz, K., Jones, B. D., Sahbaz, S., & Moqbel, M. (2021). Methodology to combine theoretical knowledge with a data-driven probabilistic graphical model. *Journal of Business Analytics*, *4*(2), 125–139.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, *30*, 5998-6008.

Wang, G., Chen, G., Zhao, H., Zhang, F., Yang, S., & Lu, T. (2021). Leveraging multisource heterogeneous data for financial risk prediction: A novel hybrid-strategy-based self-adaptive method. *MIS Quarterly*, *45*(4), 1949–1998.

Wang, G., Ma, J., & Chen, G. (2023). Attentive statement fraud detection: Distinguishing multimodal financial data with fine-grained attention. *Decision Support Systems*, *167*, 113913.

Wang, G., Ma, J., Chen, G., & Yang, Y. (2020). Financial distress prediction: Regularized sparse-based random subspace with ER aggregation rule incorporating textual disclosures. *Applied Soft Computing*, *90*, 106152.

Xia, Y., Liu, C., & Liu, N. (2017). Cost-sensitive boosted tree for loan evaluation in peer-to-peer lending. *Electronic Commerce Research and Applications*, *24*, 30–49.

Xu, J., Chen, D., Chau, M., Li, L., & Zheng, H. (2022). Peer-to-peer loan fraud detection: Constructing features from transaction data. *MIS Quarterly*, *45*(3), 1777–1792.

Yang, X., Feng, S., Wang, D., & Zhang, Y. (2021). Image-text multimodal emotion classification via multi-view attentional network. *IEEE Transactions on Multimedia*, *23*, 4014–4026.

Yang, Y., Guan, Z., Li, J., Zhao, W., Cui, J., & Wang, Q. (2021). Interpretable and efficient heterogeneous graph convolutional network. *IEEE Transactions on Knowledge and Data Engineering*, *35*(2), 1637–1650.

Yang, Y., Qin, Y., Fan, Y., & Zhang, Z. (2023). Unlocking the power of voice for financial risk prediction: A theory-driven deep learning design approach. *MIS Quarterly*, *47*(1), 63–96.

Ye, Y., & Ji, S. (2021). Sparse graph attention networks. *IEEE Transactions on Knowledge and Data Engineering*, *35*(1), 905–916.

Yin, C., Jiang, C., Jain, H. K., & Wang, Z. (2020). Evaluating the credit risk of SMEs using legal judgments. *Decision Support Systems*, *136*, 113364.

Yıldırım, M., Okay, F. Y., & Özdemir, S. (2021). Big data analytics for default prediction using graph theory. *Expert Systems with Applications*, *176*, 114840.

Zhang, Z., Wu, C., Qu, S., & Chen, X. (2022). An explainable artificial intelligence approach for financial distress prediction. *Information Processing & Management*, *59*(4), 102988.

Appendix A. Proof for the Variational Upper-Bound of Conditional Entropy

Given the two variables X and Y, let $P(Y|X)$ denote a conditional distribution and $H(Y|X)$ be the conditional entropy associated with this distribution. Then, for any $\epsilon > 0$, there exists a neural network $T_\theta(Y|X)$ such that:

$$\left|\mathrm{CE}\left(T_\theta(Y|X)\right) - H(Y|X)\right| \leq \frac{\epsilon}{2}, \qquad a.e. \tag{A1}$$

where the cross-entropy term $\mathrm{CE}\left(T_\theta(Y|X)\right) = -\mathbb{E}_{P(X,Y)}\log\left(T_\theta(Y|X)\right)$.

With the cross-entropy and KL-divergence, the conditional entropy $H(Y|X)$ can be expressed as:

$$\begin{aligned} H(Y|X) &= \mathbb{E}_{P(X,Y)}\log\frac{1}{P(Y|X)} \\ &= \mathbb{E}_{P(X,Y)}\log\frac{1}{T_\theta(Y|X)}\frac{T_\theta(Y|X)}{P(Y|X)} \\ &= \mathbb{E}_{P(X,Y)}\log\frac{1}{T_\theta(Y|X)} - D_{\mathrm{KL}}\left(P(Y|X)\|T_\theta(Y|X)\right) \end{aligned} \tag{A2}$$

Since the KL-divergence is a non-negative measure, the conditional entropy in Eq. (A2) satisfies the inequality constrain in Eq. (A3). Therefore, we can derive the variational upper-bound, i.e., $\mathrm{CE}(T_\theta(Y|X))$.

$$H(Y|X) \leq \mathrm{CE}(T_\theta(Y|X)) \tag{A3}$$

Similarly, for the three variables X, Y, and Z, the variational upper-bound of the conditional entropy $H(Z|X,Y)$ can be expressed as $\mathrm{CE}(T_\theta(Z|X,Y))$.

Appendix B. Description of Multimodal Data of NEEQ Companies.

Table B1. Description of Ratio Groups.

Category		Feature	Min	Max	Mean	S.D.
Profitability	X1	Return on assets	-2.09	1.20	0.03	0.17
	X2	Return on equity	-35.60	6.06	0.03	0.69
	X3	Net profit ratio	-676.74	250.37	-0.34	11.96
	X4	Net profit to current asset	-5.45	2.70	0.04	0.29
	X5	Net profit to fixed asset	-525.69	9,087.01	3.98	105.92
	X6	Ebit to asset	-18.95	23.48	0.04	0.40
Solvency	X7	Current ratio	0.00	4,250.78	4.10	48.59
	X8	Quick ratio	0.00	4,912.06	3.46	48.44
	X9	Asset liability ratio	0.00	197.53	0.44	2.27
	X10	Debt equity ratio	-105.16	565.26	1.29	10.65
	X11	Debt tangible equity ratio	-46,931.52	77,450.45	248.42	2,828.72
	X12	Current liability coverage	-977.18	47.72	-0.08	11.16
Development capacity	X13	Operating revenue growth rate	-1.00	18,524.68	3.12	221.03
	X14	Net profit growth rate	-5,388.92	171.35	-1.12	72.03
	X15	Assets growth rate	-1.00	27.84	0.21	0.60
	X16	Net operation cash flow growth rate	-85,640.02	84,222.50	103.52	3,796.21
	X17	Operation cash per share growth rate	-34,510.21	85,262.69	119.52	3,030.95
	X18	Equity growth rate	-91.12	111.78	0.20	2.33
Operational capabilities	X19	Inventory turning rate	0.00	133,586.30	107.98	3,103.13
	X20	Receivable turnover ratio	0.00	243,711.40	185.80	5,535.36
	X21	Accrued payable rate	0.00	9,935.29	16.56	236.86
	X22	Equity rate	0.00	91.72	1.94	3.80
	X23	Net operating cycle	-307,587.90	44,062.09	114.71	7,195.30
	X24	Working capital total rate	0.00	13,901.76	19.91	375.97
Finance structure	X25	Current asset ratio	0.01	1.00	0.71	0.22
	X26	Fixed asset ratio	0.00	0.96	0.16	0.16
	X27	Equity to fixed asset ratio	-6,804.78	37,171.16	59.77	639.05
	X28	Current liability ratio	0.06	1.00	0.93	0.14
	X29	Equity ratio	-196.53	1.00	0.56	2.27
	X30	Working capital to equity	-382.48	56.09	0.34	6.61

B.1. Description of Current Reports

A current report, also known as an 8-K Form, is a report of unscheduled material events or corporate changes at a company that could be of importance to the shareholders. Current reports provide information regarding the events disclosed, such as underlying causes, current status, and potential consequences. For

example, in the case of a company facing a judicial freeze on its equity, the company may disclose a current report to describe the reasons leading to the equity freeze (e.g., a property preservation measure), its current status (e.g., 2,977,600 shares frozen by judicial order, accounting for 14.18% of the company's total share capital), and the potential effects on the company (e.g., potential changes in controlling shareholders or actual controllers). Additionally, the current reports may also include information about the involved shareholders, such as their names, whether they hold controlling interests, the total number of shares they own, and their ownership percentage.

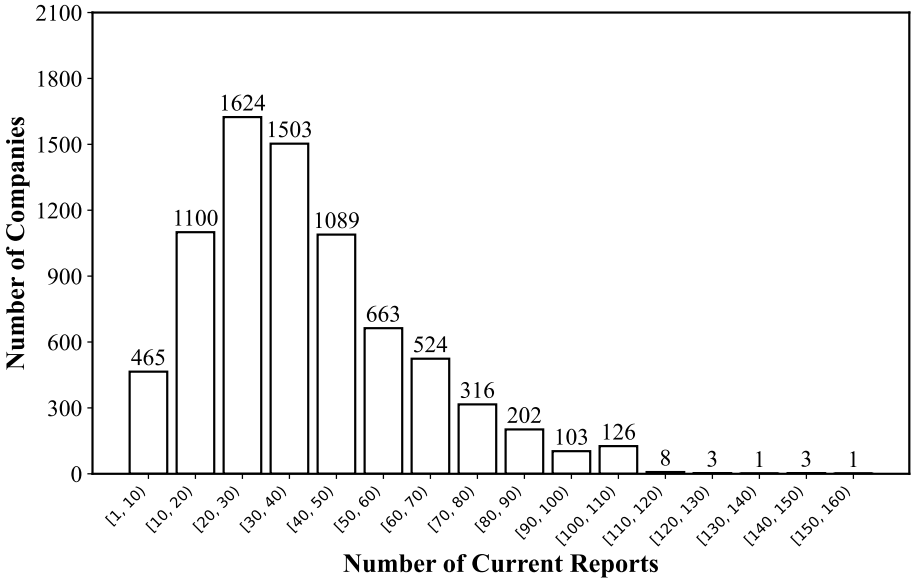

Figure B1. Distribution of Current Reports among the NEEQ Companies.

B.2. Description of Interfirm Networks

Interfirm networks refer to a type of relational network that consists of a group of mutually connected companies. An interfirm network consists of two constituent elements: nodes and edges; each node represents a company, and each edge symbolizes a relationship between two nodes, such as the sharing of directors, supervisors, or senior management. The relationships between companies indicate their interaction or association. Figure B2 illustrates an example of constructing an interfirm network. The circular nodes A–D represent four companies and the square nodes containing portraits represent personnel

in these companies, i.e., directors, supervisors, and senior management. An interfirm network is constructed, wherein two companies are linked if they share at least one director, supervisor, or senior management.

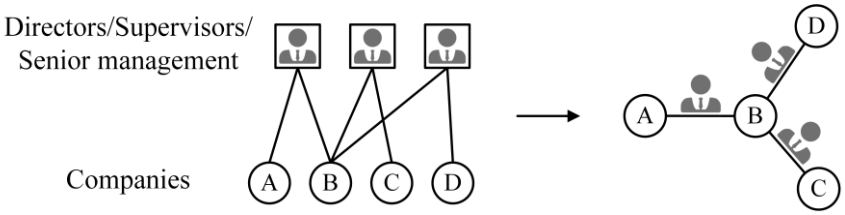

Figure B2. Example of Constructing an Interfirm Network.

Table B2. Descriptive Statistics of the Interfirm Networks.

	Directors/Supervisors/Senior Management
Average number of individuals per company	9.59
Average number of neighbor companies per company	15.63
Average number of companies per individual	3.17

Appendix C. Experiment Execution and Parameter Settings

We implemented ARDL and the benchmarked methods based on programs of tensorflow-gpu-2.0.0, keras-2.3.1, and scikit-learn-0.23.2. All experiments were run on a server with a GPU of NVIDIA RTX 2080Ti, RAM of 32 GB, and the Inter-Core-i7-9700 CPU processor with a 3.00 GHz clock speed. Table C1 and Table C2 summarize the experiment settings of key parameters for both ARDL and benchmarks.

Table C1. Parameter Settings for the ARDL.

Module	Parameter	Setting	Description
MLP layer	Numeric embedding size	128	Output dimensionality of the hidden layer of MLP
Pre-trained BERT	Text embedding size	768	Output dimensionality of the hidden layer of the pre-trained BERT
GNN layer	Graph embedding size	128	Output dimensionality of the last hidden layer of GNN
GRU layer	Representation output dimensionality of textual embeddings	128	Output dimensionality of each hidden cell of GRU for textual embeddings
Conditional entropy estimators	Number of bins	100	Dimensionality of each label for the conditional entropy estimation
Adam optimizer	Lr	0.001	Learning rate of the optimizer
Overall	Epoch	Automatically optimized by early stopping	Number of epochs to train the model
	Batch size	128	Number of samples per gradient update
	λ	Multiple settings	Conditional-entropy based regularization term in the loss function of ARDL

Table C2. Parameter Settings for the benchmarks.

Model	Parameter	Setting	Description
LR	C	{0.5, 1, 1.5, 2}	The regularization coefficient of the model
SVM	C	{1, 2, 4, 8}	The regularization coefficient of the model
XGB	Eta	{0.001, 0.01, 0.1}	The learning rate of the model
	Max_Depth	{4, 5, 6, 7, 8}	The maximum depth of a tree
LightGBM	Eta	{0.001, 0.01, 0.1}	The learning rate of the model
	Max_Depth	{4, 5, 6, 7, 8}	The maximum depth of a tree
TMN	Number of heads	{1, 2, 4, 8}	The number of heads in the multi-head attention models
	Epochs	Automatically optimized by early stopping	The number of epochs to train the model
AFF	Epochs	Automatically optimized by early stopping	The number of epochs to train the model

FGAN	Epochs	Automatically optimized by early stopping	The number of epochs to train the model
CML	Epochs	Automatically optimized by early stopping	The number of epochs to train the model
LMF	Rank	{1, 2, 4, 8, 16}	The regularization coefficient of the model controlling the output dimensionality of the low-rank representation
	Epochs	Automatically optimized by early stopping	The number of epochs to train the model
GMU	Epochs	Automatically optimized by early stopping	The number of epochs to train the model

Appendix D. Predictive Performance of ARDL versus Benchmarks on the Multimodal Dataset of NEEQ Companies

Table D1 summarizes the results of each method (mean and standard deviation) using single modality, dual modalities, and triple modalities, respectively. As for the textual modality, the input to the machine learning methods undergoes PCA processing.

Table D1. Predictive Performance of ARDL versus Benchmarks (%).

Method	Fin					Fin+Text					Fin+Net					Fin+Text+Net				
Metrics	AUC	KS	H	Precise	Recall	AUC	KS	H	Precise	Recall	AUC	KS	H	Precise	Recall	AUC	KS	H	Precise	Recall
LR	79.10 (1.91)	58.36 (1.74)	45.03 (1.86)	77.89 (1.74)	81.62 (1.84)	83.22 (2.56)	61.35 (2.86)	46.55 (2.98)	78.75 (2.76)	80.91 (2.65)	85.50 (2.72)	64.43 (2.84)	50.64 (2.59)	80.35 (2.47)	83.72 (2.64)	81.55 (2.30)	61.18 (2.24)	46.47 (2.33)	78.07 (2.38)	80.59 (2.40)
SVM	82.28 (1.85)	58.68 (1.83)	45.33 (1.57)	80.19 (1.74)	81.37 (1.76)	84.36 (2.36)	63.14 (2.54)	48.88 (2.53)	80.97 (2.04)	83.37 (2.28)	85.84 (1.93)	65.65 (2.43)	51.59 (2.71)	82.38 (1.83)	85.40 (1.98)	83.41 (2.11)	63.35 (2.48)	52.49 (2.38)	80.05 (1.54)	83.26 (1.79)
XGB	84.68 (1.50)	61.09 (1.53)	46.88 (1.76)	82.82 (2.12)	84.04 (2.12)	85.89 (2.08)	64.06 (2.11)	49.95 (2.57)	81.18 (2.59)	83.43 (2.82)	87.14 (1.98)	66.65 (2.15)	53.00 (2.21)	82.87 (2.20)	85.79 (1.96)	87.50 (1.69)	68.17 (2.28)	54.06 (1.95)	81.66 (1.90)	84.14 (2.00)
LightGBM	83.96 (1.43)	59.72 (1.78)	46.56 (1.69)	81.90 (1.96)	83.92 (1.97)	85.32 (1.89)	63.85 (2.91)	50.81 (2.50)	82.22 (2.25)	84.33 (2.82)	86.99 (2.12)	64.99 (1.98)	52.76 (2.65)	83.70 (2.04)	86.69 (2.12)	87.07 (1.79)	67.48 (1.90)	54.15 (1.67)	82.24 (1.86)	85.22 (1.86)
TMN	85.34 (1.12)	64.17 (1.21)	52.07 (1.46)	82.15 (2.01)	85.30 (2.02)	87.58 (1.60)	65.57 (1.46)	51.98 (1.84)	83.51 (2.34)	82.08 (2.83)	88.72 (1.20)	67.60 (1.70)	53.62 (1.63)	84.44 (1.96)	86.03 (2.17)	89.77 (1.36)	69.45 (1.60)	56.72 (1.41)	83.19 (1.88)	85.65 (1.91)
AFF	85.01 (1.04)	62.12 (1.48)	49.31 (1.10)	82.32 (1.67)	85.05 (1.59)	87.05 (1.77)	65.40 (1.51)	54.05 (1.75)	84.31 (2.08)	86.24 (1.93)	87.32 (1.38)	67.43 (1.41)	53.49 (1.27)	85.19 (1.70)	86.92 (1.79)	88.76 (1.76)	69.75 (1.92)	55.79 (1.70)	84.35 (1.58)	86.39 (1.93)
FGAN	85.18 (1.18)	63.30 (0.95)	50.05 (1.17)	83.20 (1.59)	85.50 (1.86)	88.29 (1.39)	66.81 (1.73)	54.40 (2.18)	85.16 (2.06)	86.80 (2.27)	88.81 (1.74)	68.81 (1.83)	56.40 (1.61)	85.78 (1.69)	87.14 (1.62)	90.94 (1.63)	71.13 (1.89)	59.90 (1.83)	84.86 (1.74)	85.99 (1.84)
CML	-	-	-	-	-	86.94 (1.62)	65.21 (1.78)	50.90 (2.03)	84.08 (2.37)	85.49 (2.01)	87.67 (1.66)	66.76 (1.85)	55.55 (1.52)	84.53 (2.31)	87.43 (2.22)	89.27 (1.33)	69.83 (2.20)	56.75 (1.47)	83.71 (2.05)	85.88 (2.20)
LMF	-	-	-	-	-	86.80 (1.78)	66.60 (2.09)	52.93 (1.68)	83.56 (2.11)	85.45 (2.21)	86.94 (1.57)	65.09 (1.91)	52.66 (1.65)	84.20 (1.99)	86.85 (1.87)	88.78 (1.91)	68.49 (2.11)	55.83 (1.55)	81.20 (1.95)	83.59 (1.96)
GMU	-	-	-	-	-	87.63 (1.51)	67.07 (1.76)	53.54 (1.50)	84.72 (2.06)	86.57 (1.88)	88.67 (1.08)	67.64 (1.70)	55.99 (1.42)	85.41 (1.69)	87.06 (1.91)	90.20 (1.66)	70.02 (2.21)	58.32 (1.68)	83.54 (1.69)	86.76 (1.78)
ARDL	**85.84** (0.80)	**65.24** (0.85)	**52.35** (0.60)	**84.50** (1.16)	**85.88** (1.01)	**89.25** (1.47)	**68.27** (1.66)	**56.10** (1.27)	**85.54** (1.63)	**87.16** (1.84)	**90.25** (1.12)	**69.30** (1.52)	**57.45** (1.24)	**86.35** (1.29)	**88.51** (1.68)	**91.76** (1.48)	**72.58** (1.58)	**62.19** (1.62)	**85.37** (2.11)	**87.39** (1.96)

Notes: “Fin” refers to financial indicators; “Text” refers to current reports; “Net” refers to interfirm networks; The best performance is in boldface.

D.1. Predictive Performance of ARDL versus Benchmarks

To exploit the potential of sentiment and readability information embedded in the current reports, we further extracted sentiment polarity and text readability features to enrich the textual input. Specifically, for sentiment analysis, we used the Chinese HowNet sentiment lexicons to capture the positive and negative sentiments expressed in the reports. Using the bag-of-words approach (unigrams) with TF-IDF, we calculated the counts and frequencies of sentiment words in each report. Subsequently, we used proportional weighting to derive the positive and negative sentiment scores for each report as the sentiment features (i.e., 2-dimension). As for measuring the Chinese financial text readability, we employed

two types of Chinese readability indexes to quantify both text capacity and flexibility. This generated 2-dimensional readability features. One of the readability features calculates the average word number in each sentence of each report. The other measures the proportion of adverbs and conjunctions in each sentence of each report using the Modern Chinese Function Words dictionary. We incorporated these newly introduced features with the semantic features, and used the integrated features as the input in terms of textual modality for both ARDL and benchmarks. Table D2 summarizes the results of each method (mean and standard deviation) using single modality, dual modalities, and triple modalities, respectively.

Table D2. Predictive Performance of ARDL versus Benchmarks (%).

Method	Fin					Fin+Text					Fin+Net					Fin+Text+Net				
Metrics	AUC	KS	H	Precise	Recall	AUC	KS	H	Precise	Recall	AUC	KS	H	Precise	Recall	AUC	KS	H	Precise	Recall
LR	79.10	58.36	45.03	77.89	81.62	83.99	62.25	47.24	79.65	81.66	85.50	64.43	50.64	80.35	83.72	82.01	61.83	46.74	78.32	81.10
	(1.91)	(1.74)	(1.86)	(1.74)	(1.84)	(2.75)	(2.74)	(2.88)	(2.77)	(2.79)	(2.72)	(2.84)	(2.59)	(2.47)	(2.64)	(2.35)	(2.13)	(2.32)	(2.33)	(2.39)
SVM	82.28	58.68	45.33	80.19	81.37	84.75	63.55	49.50	81.30	83.53	85.84	65.65	51.59	82.38	85.40	83.66	63.83	52.67	80.16	83.73
	(1.85)	(1.83)	(1.57)	(1.74)	(1.76)	(2.27)	(2.64)	(2.46)	(2.29)	(2.41)	(1.93)	(2.43)	(2.71)	(1.83)	(1.98)	(2.38)	(2.57)	(2.46)	(1.79)	(1.77)
XGB	84.68	61.09	46.88	82.82	84.04	86.21	64.96	50.11	82.04	83.90	87.14	66.65	53.00	82.87	85.79	87.70	68.56	54.70	81.78	84.50
	(1.50)	(1.53)	(1.76)	(2.12)	(2.12)	(2.28)	(2.18)	(2.49)	(2.34)	(2.91)	(1.98)	(2.15)	(2.21)	(2.20)	(1.96)	(1.55)	(2.11)	(1.86)	(2.12)	(2.17)
LightGBM	83.96	59.72	46.56	81.90	83.92	85.80	64.29	51.42	83.02	85.10	86.99	64.99	52.76	83.70	86.69	87.55	67.90	54.48	82.73	85.42
	(1.43)	(1.78)	(1.69)	(1.96)	(1.97)	(2.04)	(2.68)	(2.45)	(2.14)	(2.70)	(2.12)	(1.98)	(2.65)	(2.04)	(2.12)	(2.07)	(2.04)	(1.80)	(2.08)	(1.88)
TMN	85.34	64.17	52.07	82.15	85.30	87.61	65.68	52.11	83.78	82.16	88.72	67.60	53.62	84.44	86.03	89.79	69.49	56.72	83.31	85.76
	(1.12)	(1.21)	(1.46)	(2.01)	(2.02)	(1.64)	(1.46)	(1.85)	(2.37)	(2.85)	(1.20)	(1.70)	(1.63)	(1.96)	(2.17)	(1.39)	(1.63)	(1.36)	(1.83)	(1.89)
AFF	85.01	62.12	49.31	82.32	85.05	87.23	65.40	54.07	84.52	86.25	87.32	67.43	53.49	85.19	86.92	88.76	69.76	55.94	84.49	86.44
	(1.04)	(1.48)	(1.10)	(1.67)	(1.59)	(1.78)	(1.48)	(1.73)	(2.05)	(1.91)	(1.38)	(1.41)	(1.27)	(1.70)	(1.79)	(1.71)	(1.94)	(1.70)	(1.55)	(1.88)
FGAN	85.18	63.30	50.05	83.20	85.50	88.51	66.83	54.40	85.17	86.91	88.81	68.81	56.40	85.78	87.14	91.02	71.19	59.93	84.86	85.99
	(1.18)	(0.95)	(1.17)	(1.59)	(1.86)	(1.36)	(1.70)	(2.20)	(2.08)	(2.23)	(1.74)	(1.83)	(1.61)	(1.69)	(1.62)	(1.62)	(1.91)	(1.86)	(1.71)	(1.81)
CML	-	-	-	-	-	86.96	65.40	51.02	84.20	85.51	87.67	66.76	55.55	84.53	87.43	89.30	69.84	56.85	83.71	85.88
						(1.62)	(1.74)	(2.05)	(2.35)	(1.96)	(1.66)	(1.85)	(1.52)	(2.31)	(2.22)	(1.37)	(2.19)	(1.45)	(2.01)	(2.20)
LMF	-	-	-	-	-	86.80	66.74	53.02	83.83	85.59	86.94	65.09	52.66	84.20	86.85	88.80	68.56	56.01	81.20	83.65
						(1.79)	(2.08)	(1.72)	(2.12)	(2.24)	(1.57)	(1.91)	(1.65)	(1.99)	(1.87)	(1.93)	(2.10)	(1.59)	(1.95)	(1.96)
GMU	-	-	-	-	-	87.64	67.15	53.67	84.75	86.59	88.67	67.64	55.99	85.41	87.06	90.24	70.08	58.35	83.74	86.82
						(1.46)	(1.75)	(1.49)	(2.07)	(1.89)	(1.08)	(1.70)	(1.42)	(1.69)	(1.91)	(1.66)	(2.25)	(1.69)	(1.67)	(1.78)
ARDL	**85.84**	**65.24**	**52.35**	**84.50**	**85.88**	**89.25**	**68.27**	**56.13**	**85.56**	**87.28**	**90.25**	**69.30**	**57.45**	**86.35**	**88.51**	**91.76**	**72.59**	**62.21**	**85.39**	**87.51**
	(0.80)	(0.85)	(0.60)	(1.16)	(1.01)	(1.50)	(1.63)	(1.26)	(1.67)	(1.80)	(1.12)	(1.52)	(1.24)	(1.29)	(1.68)	(1.46)	(1.54)	(1.62)	(2.11)	(1.92)

Notes: “Fin” refers to financial indicators; “Text” refers to current reports; “Net” refers to interfirm networks; The textual input comprises semantic, sentiment, and readability features. The best performance is in boldface.

Appendix E. Description of Multimodal Data of Listed Companies

We collected 5,206 listed companies in the Shanghai Stock Exchange and Shenzhen Stock Exchange from 2019 to 2022. We used multimodal data of each company by end of 2019 to predict financial distress of the year 2022. Based on ST, there were 118 financial distressed companies (accounting for 2.26% of the total) and 5,088 normal operated companies. For predictors of financial distress, we collected data of the aforementioned three modalities, i.e., financial ratios, current reports, and interfirm networks.

E.1. Description of Financial Ratios

We collected 30 financial ratios in the 2019 annual report of each company from the CSMAR dataset. We divided the financial ratios into five groups based on the aspect of financial conditions they reflect, i.e., profitability, solvency, development capacity, operational capabilities, and finance structure. Table E1 summarizes the descriptive statistics of these financial ratios.

Table E1. Description of Ratio Groups.

Category		Feature	Min	Max	Mean	S.D.
Profitability	X1	Return on assets	-1.46	1.72	0.04	0.11
	X2	Return on equity	-80.29	6.23	0.03	1.30
	X3	Net profit ratio	-1,724.40	3,477,431.34	843.54	49,732.37
	X4	Net profit to current asset	-8.70	10.54	0.07	0.55
	X5	Net profit to fixed asset	-5,825.27	24,217.53	17.45	614.14
	X6	Ebit to asset	-0.65	31.87	0.05	0.55
Solvency	X7	Current ratio	0.00	80.66	2.50	3.03
	X8	Quick ratio	0.00	52.14	2.03	2.75
	X9	Asset liability ratio	0.01	2.12	0.42	0.21
	X10	Debt equity ratio	-153.09	102.98	3.14	6.32
	X11	Debt tangible equity ratio	-5,062.53	104,835.85	210.79	1,611.81
	X12	Current liability coverage	-16.12	133.67	0.24	1.90
Development capacity	X13	Operating revenue growth rate	-13.09	2,637.55	2.82	48.23
	X14	Net profit growth rate	-2,823.20	60.05	-1.60	40.80
	X15	Assets growth rate	-0.93	77.70	0.19	1.32
	X16	Net operation cash flow growth rate	-211,427.84	253,290.13	260.83	7,947.75
	X17	Operation cash per share growth rate	-213,946.92	135,166.67	344.07	6,713.55
	X18	Equity growth rate	-250.93	259.82	0.23	7.86
Operational capabilities	X19	Inventory turning rate	0.00	380,072.47	158.14	5,492.41
	X20	Receivable turnover ratio	0.00	4,332,756.29	909.74	60,092.26
	X21	Accrued payable rate	0.00	61,277.37	36.24	1,082.41
	X22	Equity rate	0.00	119.75	4.92	5.47

	X23	Net operating cycle	-578,446.36	114,707.73	303.98	1,970.54
	X24	Working capital total rate	0.00	4,615.26	102.39	169.79
Finance structure	X25	Current asset ratio	0.00	1.00	0.59	0.22
	X26	Fixed asset ratio	0.00	0.95	0.19	0.15
	X27	Equity to fixed asset ratio	-715.51	7,208.71	19.52	133.65
	X28	Current liability ratio	0.00	1.01	0.82	0.20
	X29	Equity ratio	-1.12	0.99	0.58	0.21
	X30	Working capital to equity	-3.10	2.42	0.96	0.11

E.2. Description of Current Reports

We collected all the current reports disclosed by each company from the official disclosure platform[4] of listed companies in 2019, resulting in 244,226 current reports. The mean and standard deviation of the number of current reports per company are 46.912 and 45.451, respectively. Similarly, we utilized Tesseract OCR technology to extract text from images in the current reports. Figure E1 illustrates the distribution of current reports among companies.

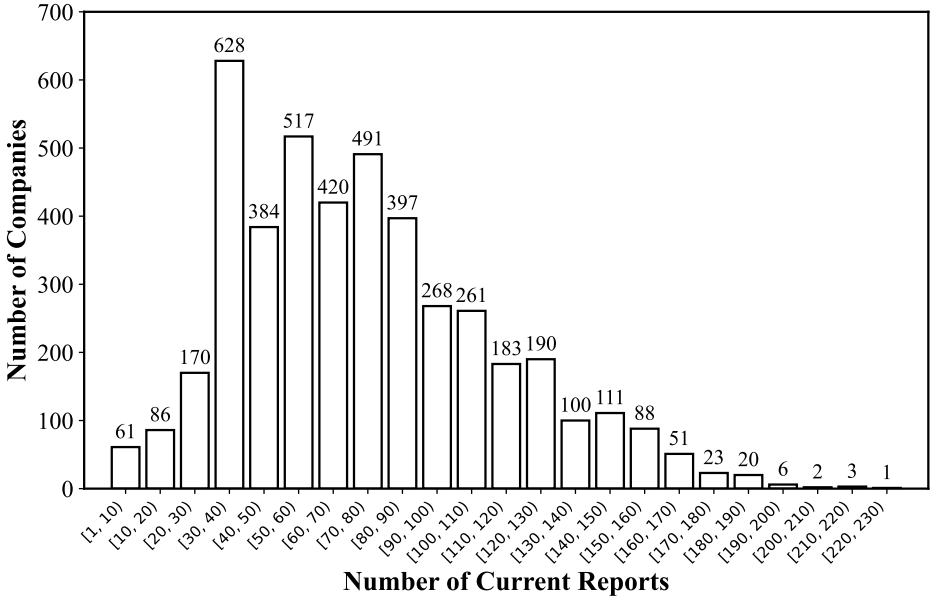

Figure E1. Distribution of Current Reports among the Listed Companies.

[4] The platform (http://eid.csrc.gov.cn/) offers publicly available information about listed companies in China.

E.3. Description of Interfirm Networks

We collected information regarding the neighbor companies of the 5,206 listed companies in our dataset from Qichacha and built interfirm networks. Based on the directors, supervisors, and senior management (DSS) relationships between companies, we identified 191,065 neighbor companies of the original 5,206 companies. Table E2 summarizes the descriptive statistics of the DSS networks. For attributes to characterize each node (i.e., company) in DSS networks, we selected demographic attributes (i.e., company age, district, industry type, registered capital, number of patents, and percentage of insider ownership) and risk event attributes (i.e., administrative penalties, equity pledges, and loan disputes). The risk event attributes were set to 1 if the company was involved in the corresponding risk events in 2019, and 0 otherwise.

Table E2. Descriptive Statistics of the Interfirm Networks.

	Directors/Supervisors/Senior Management
Average number of individuals per company	16.098
Average number of neighbor companies per company	36.701
Average number of companies per individual	10.882

Appendix F. Predictive Performance of ARDL versus Benchmarks on the Multimodal Dataset of Listed Companies

Table F1. Predictive Performance of ARDL versus Benchmarks (%).

Method	Fin					Fin+Text					Fin+Net					Fin+Text+Net				
Metrics	AUC	KS	H	Precise	Recall	AUC	KS	H	Precise	Recall	AUC	KS	H	Precise	Recall	AUC	KS	H	Precise	Recall
LR	83.91 (2.16)	60.95 (1.85)	47.47 (1.68)	81.13 (1.71)	83.75 (1.82)	85.46 (2.83)	63.35 (2.64)	48.93 (2.59)	82.49 (2.75)	84.31 (2.37)	87.52 (2.81)	67.64 (2.36)	53.24 (2.60)	83.14 (2.56)	85.76 (2.28)	87.77 (2.32)	68.88 (1.74)	53.71 (2.08)	83.72 (2.26)	85.93 (2.56)
SVM	84.58 (1.91)	61.93 (1.77)	47.77 (1.86)	82.79 (1.53)	83.95 (1.98)	85.92 (2.16)	62.19 (2.26)	50.89 (2.67)	84.77 (1.82)	84.83 (2.37)	87.87 (2.01)	68.25 (1.95)	53.48 (2.49)	84.63 (1.98)	87.46 (1.93)	88.90 (2.18)	68.65 (2.68)	55.36 (2.47)	85.42 (1.95)	88.09 (1.58)
XGB	86.65 (1.47)	63.21 (1.24)	49.03 (1.61)	85.55 (2.19)	86.16 (1.97)	87.91 (2.04)	65.91 (1.87)	51.03 (1.99)	86.91 (2.30)	86.51 (2.53)	89.31 (1.75)	68.48 (1.75)	55.56 (2.17)	87.88 (1.75)	86.99 (1.88)	90.17 (2.22)	71.36 (1.89)	57.06 (1.36)	88.97 (2.13)	89.96 (1.98)
LightGBM	86.78 (1.33)	62.55 (1.39)	48.99 (1.26)	84.34 (2.07)	86.74 (1.51)	87.05 (2.17)	66.12 (2.85)	52.50 (2.42)	85.96 (2.36)	86.85 (2.54)	88.95 (1.55)	67.46 (1.87)	55.66 (2.46)	85.32 (2.19)	88.55 (2.09)	89.96 (1.72)	70.14 (2.12)	56.94 (2.31)	88.51 (1.57)	89.31 (2.05)
TMN	85.22 (1.28)	63.42 (1.36)	51.68 (1.25)	85.34 (1.97)	87.15 (1.89)	89.12 (1.50)	66.24 (1.02)	53.56 (1.40)	86.44 (2.03)	88.26 (2.82)	91.07 (1.27)	70.12 (1.39)	56.66 (1.17)	86.26 (1.76)	89.82 (1.71)	92.61 (1.19)	72.07 (1.73)	58.97 (1.62)	89.15 (1.95)	90.45 (1.86)
AFF	87.06 (1.39)	65.32 (1.52)	50.36 (0.94)	85.04 (1.17)	88.01 (1.28)	89.01 (1.81)	66.64 (1.38)	54.57 (1.88)	85.99 (2.29)	88.36 (1.84)	89.99 (1.47)	69.83 (1.44)	56.12 (1.18)	87.51 (1.53)	88.65 (2.14)	91.52 (1.35)	72.96 (1.61)	58.24 (1.25)	88.37 (1.55)	88.31 (1.47)
FGAN	87.31 (1.15)	65.47 (1.31)	53.25 (0.83)	85.93 (1.62)	87.74 (1.86)	89.63 (1.57)	67.33 (1.89)	55.18 (1.96)	86.96 (1.91)	87.99 (1.95)	91.55 (1.46)	71.27 (1.58)	58.84 (1.44)	88.36 (1.74)	90.01 (1.79)	92.43 (1.74)	73.75 (1.48)	63.32 (1.71)	88.77 (1.26)	90.28 (1.55)
CML	-	-	-	-	-	88.77 (1.61)	65.82 (1.97)	52.54 (2.08)	84.82 (2.19)	86.88 (2.14)	89.77 (1.72)	69.07 (1.73)	57.42 (1.61)	86.65 (2.21)	89.61 (2.09)	91.14 (1.47)	71.73 (2.22)	58.66 (1.25)	87.95 (2.09)	88.35 (2.22)
LMF	-	-	-	-	-	86.79 (1.55)	64.33 (1.67)	51.43 (1.78)	84.81 (2.22)	87.26 (2.36)	90.18 (1.56)	68.13 (1.99)	56.92 (1.85)	86.57 (1.71)	88.27 (1.48)	90.61 (1.64)	71.51 (2.05)	57.74 (1.52)	86.44 (1.87)	86.44 (1.93)
GMU	-	-	-	-	-	88.89 (1.69)	68.84 (1.46)	54.83 (1.56)	85.78 (2.24)	88.22 (1.43)	91.31 (1.32)	70.04 (1.31)	59.10 (1.26)	88.38 (1.59)	89.43 (1.51)	92.36 (1.16)	72.75 (2.18)	60.62 (1.65)	89.78 (1.87)	89.03 (1.47)
ARDL	**88.07** (0.99)	**68.02** (0.86)	**54.53** (0.95)	**86.31** (1.05)	**88.68** (1.11)	**90.39** (1.52)	**69.05** (1.77)	**56.68** (1.39)	**87.10** (1.42)	**88.81** (1.39)	**92.12** (1.41)	**72.46** (1.68)	**59.61** (1.32)	**89.97** (1.16)	**90.88** (1.62)	**94.25** (1.57)	**75.62** (1.32)	**64.01** (1.62)	**91.28** (1.99)	**90.68** (1.89)

Notes: “Fin” refers to financial indicators; “Text” refers to current reports; “Net” refers to interfirm networks. The best performance is in boldface.

Table F2. Representation Performance of ARDL versus Benchmarks (%).

Method	LR					SVM					XGB					LightGBM				
Metrics	AUC	KS	H	Precise	Recall	AUC	KS	H	Precise	Recall	AUC	KS	H	Precise	Recall	AUC	KS	H	Precise	Recall
TMN	91.83 (1.71)	71.37 (1.68)	61.03 (1.85)	89.47 (1.79)	89.16 (1.55)	92.23 (1.52)	70.84 (1.67)	61.23 (1.84)	89.69 (1.78)	88.01 (1.64)	93.76 (1.61)	75.22 (1.98)	64.57 (1.53)	89.71 (1.69)	89.51 (1.56)	93.07 (1.60)	72.68 (2.06)	62.48 (1.52)	85.66 (1.59)	86.78 (1.64)
AFF	91.09 (1.94)	71.07 (1.83)	59.93 (1.43)	89.69 (1.69)	88.17 (1.73)	91.19 (1.53)	71.51 (1.91)	58.24 (1.65)	88.08 (1.64)	88.42 (1.97)	92.82 (1.73)	74.71 (2.13)	62.02 (1.56)	88.85 (1.79)	87.35 (1.42)	92.23 (1.67)	72.12 (1.95)	61.06 (1.73)	86.34 (1.22)	86.54 (1.63)
FGAN	92.28 (1.55)	72.45 (1.78)	62.45 (1.54)	88.85 (1.64)	89.31 (1.29)	92.45 (1.57)	72.83 (2.05)	62.75 (1.72)	89.46 (1.77)	88.76 (1.58)	92.49 (1.79)	76.67 (1.65)	64.84 (1.46)	91.25 (1.44)	88.91 (1.57)	94.01 (1.82)	73.87 (1.78)	63.13 (1.42)	87.39 (1.90)	88.08 (1.73)
CML	91.46 (1.43)	69.79 (1.91)	59.08 (1.86)	87.26 (1.48)	87.03 (1.52)	89.86 (1.90)	69.17 (1.97)	57.95 (1.88)	87.63 (1.55)	87.46 (1.77)	91.80 (1.67)	73.74 (1.84)	61.53 (1.92)	88.62 (1.84)	87.44 (1.54)	93.05 (1.65)	71.82 (2.08)	60.73 (1.32)	86.32 (1.75)	86.92 (1.38)
LMF	90.31 (1.73)	69.60 (1.99)	57.21 (1.58)	88.63 (1.78)	86.74 (1.61)	89.86 (1.71)	68.81 (2.16)	57.21 (1.61)	86.01 (1.68)	87.35 (2.15)	91.24 (1.85)	72.76 (1.97)	60.64 (1.55)	88.89 (1.89)	86.18 (1.93)	92.70 (2.12)	69.26 (2.35)	58.75 (1.88)	85.94 (1.83)	85.23 (1.58)
GMU	92.89 (1.51)	73.05 (1.59)	62.75 (1.51)	88.01 (1.37)	88.75 (1.49)	91.16 (1.57)	73.62 (2.17)	63.33 (1.78)	89.16 (1.65)	87.38 (1.56)	93.78 (1.51)	76.26 (1.61)	64.73 (1.46)	92.28 (1.57)	88.66 (1.83)	92.32 (1.47)	72.83 (1.49)	63.97 (1.31)	86.56 (1.36)	88.51 (1.42)
ARDL	**93.74** (1.42)	**73.71** (1.36)	**63.48** (1.13)	**90.85** (1.29)	**90.56** (1.34)	**93.79** (1.38)	**75.68** (1.59)	**65.15** (1.32)	**90.63** (1.34)	**89.07** (1.46)	**95.96** (1.26)	**77.42** (1.45)	**66.76** (1.29)	**93.69** (1.16)	**91.92** (1.51)	**94.65** (1.29)	**74.79** (1.66)	**65.15** (1.29)	**87.79** (1.12)	**89.54** (1.27)

Note: The results of representation performance of each method using trimodal data; The best performance is in boldface.

Appendix G. Description of Multi-View Data of Listed Companies

We collected 5,206 listed companies in the Shanghai Stock Exchange and Shenzhen Stock Exchange from 2019 to 2022. We used multi-view data of each company by end of 2019 to predict financial distress of the year 2022. For predictors of financial distress, we collected five views of data from the CSMAR and CNRDS (Chinese Research Data Services) database, including financial indicator view, annual report view, stock forum view, legal judgment view, and financial news view.

G.1. Description of Financial Indicator View

For the financial indicator view, we selected widely used financial ratios from the aspects of solvency, profitability, operating capacity, development capacity, and cash flow, resulting in 25 features. Table G1 summarizes the features in the financial indicator view.

Table G1. Description of Financial Indicator View.

No.	Feature	Min	Max	Mean	S.D.
1	Current ratio	0.12	47.53	2.45	2.47
2	Quick ratio	0.02	41.58	1.97	2.26
3	Interest coverage ratio	-2,750.60	104,760.40	98.68	1,684.93
4	Net cash flow from operating activities	-5.67	6.80	0.24	0.49
5	Debt to asset ratio	0.03	1.50	0.39	0.18
6	Debt to equity ratio	-2.83	70.73	0.95	2.02
7	Capital accumulation rate	-0.94	37.76	0.33	0.96
8	Total assets growth rate	-0.60	13.35	0.27	0.54
9	Net profit growth rate	-48.88	455.80	0.72	9.24
10	Administrative expenses growth rate	-0.87	10.21	0.20	0.47
11	Stockholder’s equity growth ratio	-0.91	153.58	0.39	2.52
12	Accounts receivable turnover ratio	0.34	47,889,094.68	11,306.84	742,128.60
13	Inventory turnover ratio	0.00	57,980.34	74.47	1,419.03
14	Current assets turnover ratio	0.00	12.70	1.17	0.90
15	Total assets turnover ratio	0.00	10.28	0.62	0.52
16	Equity turnover ratio	0.00	65.58	1.34	2.37
17	Net profit cash coverage	-1.21	0.54	0.08	0.07
18	Net cash content of operating profit	-1.22	0.50	0.06	0.07
19	Cash to total assets recovery ratio	-16.89	1.32	0.09	0.34
20	Price to cash flow ratio	-0.44	0.98	0.32	0.16
21	Return on assets ratio	-331.62	3.78	-0.02	5.45
22	Net profit margin on total assets	-596.25	594.68	1.06	18.53
23	Return on equity	-466.47	188.46	0.62	9.61
24	Gross margin ratio	-0.68	0.44	0.05	0.08
25	Net profit margin	-90.97	7,032.79	2.84	110.19

G.2. Description of Annual Reports View

For the annual report view, we constructed 13 features using the text of annual reports and management discussion and analysis (MD&A). We focused on statistical (e.g., number of words) and sentiment features of the texts. The Loughran-McDonald (LM) and NTU sentiment lexicons were used for annual reports and the Loughran and McDonald sentiment lexicon was used for MD&A. Table G2 summarizes the features in the annual report view.

Table G2. Description of Annual Reports View.

No.	Feature	Min	Max	Mean	S.D.
1	Number of characters in annual report	53,015.43	906,433.10	165,254.70	40,066.79
2	Number of words in annual report	15,204.01	129,857.30	49,138.35	10,577.57
3	Number of sentences in annual report	232.55	22,447.36	1,096.58	459.17
4	Number of positive words (LM) in annual report	1,031.95	8,368.96	3,598.02	747.15
5	Number of negative words (LM) in annual report	781.18	10,053.80	3,538.54	706.80
6	Sentiment tone (LM) of annual report	1,151.72	11,633.34	3,556.26	799.16
7	Number of positive words (NTU) in annual report	294.44	3,544.82	1,466.56	278.27
8	Number of negative words (NTU) in annual report	-0.20	0.23	0.00	0.05
9	Sentiment tone (NTU) of annual report	0.13	0.64	0.40	0.08
10	Number of sentences in MD&A	64.44	1,613.41	423.22	183.65
11	Number of words in MD&A	41.40	863.63	160.98	76.59
12	Number of positive words in MD&A	16.20	792.79	135.07	67.04
13	Number of negative words in MD&A	1,251.41	80,699.02	8,228.39	4,091.14

G.3. Description of Stock Forum View

For the stock forum view, we constructed 20 features based on posts and comments on Guba, one of the largest online stock forums in China, from the first quarter (Q1) to the fourth quarter (Q4) in 2019. Online stock forum is one of the major platforms for individual investors to communicate with each other and valuable information may be conveyed during communication in form of posts and comments. The interaction and dissemination of information, as well as the emergence and fluctuation of sentiment, may reflect the business and financial conditions of a company to some extent, and thus may be valuable in predicting financial distress. We therefore selected the stock forum information as another data view to complement the financial indicator view. Table G3 summarizes the features in the stock forum view.

Table G3. Description of Stock Forum View.

No.	Feature	Min	Max	Mean	S.D.
1	Number of posts in Q1	0.00	30,826.02	1,660.58	2,099.85

2	Number of positive posts in Q1	0.00	7,369.55	515.51	558.82
3	Number of negative posts in Q1	0.00	6,492.71	403.51	465.73
4	Number of reads of posts in Q1	0.00	63,320,303.95	1,853,983.74	2,796,212.74
5	Number of comments of posts in Q1	0.00	58,965.72	2,371.49	4,052.84
6	Number of posts in Q2	0.00	26,122.97	1,894.01	2,132.96
7	Number of positive posts in Q2	0.00	8,552.96	587.77	583.46
8	Number of negative posts in Q2	0.00	6,156.20	464.26	490.53
9	Number of reads of posts in Q2	0.00	52,320,307.08	3,150,438.77	3,694,050.21
10	Number of comments of posts in Q2	0.00	59,012.84	2,854.44	4,510.25
11	Number of posts in Q3	0.00	64,856.66	2,689.16	2,542.14
12	Number of positive posts in Q3	0.00	17,950.39	906.10	754.38
13	Number of negative posts in Q3	0.00	14,080.44	613.71	591.72
14	Number of reads of posts in Q3	0.00	109,451,109.90	2,980,464.89	4,070,931.37
15	Number of comments of posts in Q3	0.00	143,107.86	2,654.35	5,051.47
16	Number of posts in Q4	19.64	43,822.46	1,908.11	2,139.44
17	Number of positive posts in Q4	5.44	12,492.79	552.88	565.59
18	Number of negative posts in Q4	1.84	8,851.85	432.84	487.31
19	Number of reads of posts in Q4	14,093.64	114,699,384.46	3,240,467.11	4,374,223.53
20	Number of comments of posts in Q4	3.96	125,939.52	2,568.72	5,277.42

G.4. Description of Legal Judgment View

For the legal judgment view, we constructed 12 features based on legal judgments of each company in 2019. Legal judgment information helps to evaluate the credit risk of a company. Each legal judgment reflects the dispute of a company in the process of production and operation, and adverse outcomes (e.g., compensation) may aggravate financial risk. In this regard, legal judgments may also provide complementary information to financial information for financial distress prediction. Table G4 summarizes the features in the legal judgment view.

Table G4. Description of Legal Judgment View.

No.	Feature	Min	Max	Mean	S.D.
1	Number of judgments in role of plaintiff	0.00	69.74	1.90	4.41
2	Number of civil litigations in role of plaintiff	0.00	20.94	0.15	1.02
3	Number of criminal litigations in role of plaintiff	0.00	58.19	1.62	3.94
4	Number of arbitrations in role of plaintiff	0.00	11.07	0.17	0.68
5	Amount involved in judgments in role of plaintiff	0.00	508,277.92	6,167.75	29,362.66
6	Number of loan dispute cases in role of plaintiff	0.00	77.11	2.22	4.81
7	Number of judgments in role of defendants	0.00	26.59	0.08	0.96
8	Number of civil litigations in role of defendants	0.00	76.12	1.99	4.32
9	Number of criminal litigations in role of defendants	0.00	2.77	0.03	0.29
10	Number of arbitrations in role of defendants	0.00	369,096.82	6,768.60	25,188.97
11	Amount involved in judgments in role of defendants	0.00	68.95	1.82	4.87
12	Number of loan dispute cases in role of defendants	0.00	22.08	0.14	1.01

G.5. Description of Financial News View

For the financial news view, we constructed 18 features based on the financial news and search trends related to each company from Q1 to Q4 in 2019. Financial news exposes positive and negative events related to a company and search trends reflect the degree of attention of users (e.g., investors). Both these two types of information may be beneficial for reflecting business condition and development potential of a company, thus may contribute to financial distress prediction. Table G5 summarizes the features in the financial news view.

Table G5. Description of Financial News View.

No.	Feature	Min	Max	Mean	S.D.
1	Number of negative news in Q1	0.00	1,201.09	7.69	29.09
2	Number of positive news in Q1	0.00	804.55	9.91	28.62
3	Number of negative news in Q2	0.00	1,499.58	9.78	34.65
4	Number of positive news in Q2	0.00	1,045.84	12.29	37.13
5	Number of negative news in Q3	0.00	2,732.78	29.61	93.36
6	Number of positive news in Q3	0.00	2,893.13	50.10	141.18
7	Number of negative news in Q4	0.00	1,804.04	30.57	102.96
8	Number of positive news in Q4	0.00	4,122.03	54.26	189.55
9	Number of negative news related to executives	0.00	1,397.73	11.97	41.17
10	Number of positive news related to executives	0.00	1,797.46	11.58	56.13
11	Index of searches (stock code) in Q1	0.00	260,616.29	28,497.32	20,932.04
12	Index of searches (all related keywords) in Q1	0.00	2,031,094.99	96,619.29	101,766.97
13	Index of searches (stock code) in Q2	0.00	321,050.27	26,445.86	21,463.11
14	Index of searches (all related keywords) in Q2	0.00	1,811,917.57	95,292.83	104,879.97
15	Index of searches (stock code) in Q3	0.00	241,874.42	27,440.94	19,631.34
16	Index of searches (all related keywords) in Q3	0.00	1,963,948.50	97,580.04	112,470.64
17	Index of searches (stock code) in Q4	0.00	458,537.25	23,720.80	17,714.82
18	Index of searches (all related keywords) in Q4	0.00	1,946,007.42	82,812.14	98,088.39

Appendix H. Predictive and Representation Performance of General Variant of ARDL versus Benchmarks on the Multi-View Dataset of Listed Companies

Table 5. Predictive Performance of General Variant of ARDL versus Benchmarks (%).

Method	Finance Indicator View					Non-Financial Indicator Views					All Views				
Metrics	AUC	KS	H	Precise	Recall	AUC	KS	H	Precise	Recall	AUC	KS	H	Precise	Recall
LR	76.79 (2.43)	59.39 (2.32)	50.04 (2.33)	74.83 (2.22)	76.19 (1.67)	77.07 (2.56)	58.44 (2.11)	49.19 (1.98)	76.68 (2.20)	76.82 (1.51)	81.42 (2.09)	63.48 (1.88)	49.74 (1.78)	76.77 (2.78)	78.96 (1.55)
SVM	75.63 (2.16)	60.03 (1.92)	50.53 (1.92)	74.64 (1.99)	75.47 (2.15)	77.52 (2.29)	61.89 (1.77)	51.47 (1.42)	77.83 (1.95)	78.85 (1.78)	81.46 (1.93)	64.28 (1.59)	48.72 (1.86)	78.21 (1.65)	79.58 (1.84)
XGB	80.01 (2.13)	62.96 (2.08)	53.78 (1.99)	79.22 (2.06)	81.27 (1.65)	81.91 (1.63)	64.51 (1.57)	52.89 (1.97)	78.05 (1.84)	78.03 (1.29)	83.33 (1.83)	65.45 (1.94)	52.15 (1.81)	78.25 (1.71)	80.04 (1.59)
LightGBM	81.32 (2.40)	63.05 (2.02)	52.72 (2.12)	80.79 (2.18)	80.39 (1.42)	81.01 (1.65)	63.19 (1.51)	52.75 (1.93)	79.42 (1.91)	79.45 (1.31)	82.62 (2.14)	65.31 (1.98)	51.59 (2.08)	79.43 (2.16)	81.57 (1.26)
TMNv	82.45 (1.62)	63.54 (1.81)	53.65 (1.54)	81.01 (1.47)	82.21 (1.84)	82.87 (1.67)	63.92 (1.80)	53.64 (1.45)	80.64 (1.52)	80.66 (1.82)	84.64 (1.46)	66.89 (1.78)	54.06 (1.32)	81.26 (1.28)	82.19 (1.75)
AFF	81.68 (2.09)	62.79 (1.63)	51.66 (1.93)	81.71 (1.95)	81.11 (1.74)	82.60 (2.16)	63.65 (1.46)	51.75 (1.62)	79.74 (1.86)	80.48 (1.61)	84.46 (1.82)	66.06 (1.52)	52.73 (1.67)	79.99 (1.85)	81.28 (1.72)
FGANv	83.07 (1.52)	64.02 (1.66)	55.12 (1.76)	82.75 (1.39)	83.83 (1.33)	83.43 (1.63)	65.65 (1.46)	53.37 (1.42)	81.26 (1.28)	82.96 (1.22)	86.35 (1.32)	66.79 (1.37)	53.40 (1.53)	81.93 (1.15)	83.97 (1.09)
CML	-	-	-	-	-	81.67 (2.52)	63.11 (2.04)	53.17 (1.52)	79.17 (1.96)	80.42 (2.05)	84.35 (1.64)	66.11 (1.09)	52.11 (1.77)	81.35 (1.37)	81.95 (1.36)
LMF	-	-	-	-	-	79.65 (2.23)	62.96 (2.51)	52.38 (1.85)	78.31 (1.97)	78.55 (2.36)	83.66 (1.63)	64.04 (1.94)	50.88 (1.51)	78.69 (1.57)	79.84 (1.85)
GMU	-	-	-	-	-	82.76 (1.79)	64.01 (1.43)	53.09 (1.52)	80.07 (1.43)	81.32 (1.91)	86.22 (1.77)	67.68 (1.58)	54.98 (1.77)	82.97 (1.69)	83.68 (1.93)
ARDLv	**83.76** (1.21)	**65.42** (1.55)	**57.07** (1.97)	**82.94** (1.12)	**83.88** (1.85)	**84.96** (1.69)	**66.95** (1.12)	**54.34** (1.71)	**83.44** (1.39)	**84.05** (1.64)	**86.82** (1.12)	**68.15** (1.33)	**55.95** (1.83)	**85.45** (0.98)	**86.45** (1.49)

Notes: TMNv, FGANv, and ARDLv refer to the variants of the TMN, FGAN, and ARDL methods, each of which drops the representation learning modules; The best performance is in boldface.

Table 6. Representation Performance of General Variant of ARDL versus Benchmarks (%).

Method	LR					SVM					XGB					LightGBM				
Metrics	AUC	KS	H	Precise	Recall	AUC	KS	H	Precise	Recall	AUC	KS	H	Precise	Recall	AUC	KS	H	Precise	Recall
TMNv	84.15 (1.69)	64.01 (2.08)	54.82 (1.92)	77.52 (1.83)	78.63 (1.92)	84.16 (1.83)	63.86 (1.84)	53.41 (1.36)	77.35 (1.69)	77.89 (2.02)	85.31 (1.82)	65.61 (1.67)	54.19 (1.68)	77.84 (1.92)	78.75 (1.52)	84.38 (1.49)	65.83 (1.68)	55.23 (1.55)	78.11 (1.87)	79.31 (1.79)
AFF	83.95 (1.77)	64.17 (1.99)	54.65 (1.61)	78.09 (1.72)	78.71 (1.61)	84.06 (1.86)	64.09 (2.45)	52.85 (1.79)	77.45 (2.05)	78.38 (1.64)	85.54 (1.68)	66.65 (1.64)	56.26 (1.75)	79.13 (1.61)	78.74 (1.59)	84.81 (1.85)	66.46 (2.09)	55.75 (1.88)	79.51 (1.73)	79.34 (1.51)
FGANv	85.17 (1.99)	66.16 (1.58)	56.81 (1.77)	79.31 (1.48)	80.01 (1.31)	84.96 (1.65)	66.21 (1.74)	56.74 (1.76)	78.81 (1.73)	79.46 (1.89)	86.26 (1.71)	67.38 (2.06)	58.78 (1.61)	79.98 (1.73)	79.74 (1.64)	85.89 (1.63)	67.81 (1.82)	58.92 (1.61)	80.83 (1.84)	81.27 (1.64)
CML	84.88 (2.13)	64.28 (1.84)	53.64 (1.75)	78.45 (2.05)	79.53 (1.87)	83.19 (1.61)	64.74 (2.23)	52.83 (1.82)	78.27 (1.72)	78.85 (1.63)	85.49 (1.88)	65.88 (1.97)	53.41 (1.89)	80.03 (2.31)	79.53 (1.42)	85.61 (1.73)	66.52 (1.96)	54.09 (1.82)	80.09 (2.06)	80.39 (1.94)
LMF	83.65 (2.79)	62.26 (2.63)	52.24 (1.86)	77.72 (1.99)	77.31 (2.01)	82.28 (2.36)	61.68 (2.71)	52.37 (2.32)	76.15 (2.33)	78.04 (2.19)	84.34 (2.08)	63.04 (2.45)	53.06 (1.84)	78.02 (1.92)	77.69 (2.06)	84.37 (1.99)	63.88 (2.28)	53.23 (2.13)	78.73 (2.10)	78.95 (2.01)
GMU	84.28 (1.67)	65.79 (1.87)	56.81 (1.57)	78.44 (1.63)	79.13 (1.56)	84.93 (1.72)	66.18 (2.22)	56.85 (1.86)	78.79 (1.67)	78.44 (1.72)	86.74 (1.78)	66.82 (1.75)	58.50 (1.63)	78.59 (1.84)	79.36 (1.44)	85.28 (1.76)	67.06 (1.89)	57.62 (1.71)	80.72 (1.93)	80.59 (1.71)
ARDLv	**85.93** (1.68)	**67.19** (1.54)	**57.92** (1.69)	**80.16** (1.33)	**81.19** (1.49)	**86.32** (1.67)	**67.41** (1.89)	**58.94** (1.78)	**79.78** (1.32)	**80.03** (1.63)	**87.28** (1.67)	**69.58** (1.53)	**60.83** (1.61)	**81.11** (1.59)	**82.87** (1.35)	**86.77** (1.38)	**68.08** (1.79)	**60.07** (1.59)	**81.49** (1.28)	**82.95** (1.37)

Note: The results of representation performance of each method using all data views; The best performance is in boldface.